

The Stability of Matter: From Atoms to Stars

Selecta of Elliott H. Lieb

Edited by W. Thirring

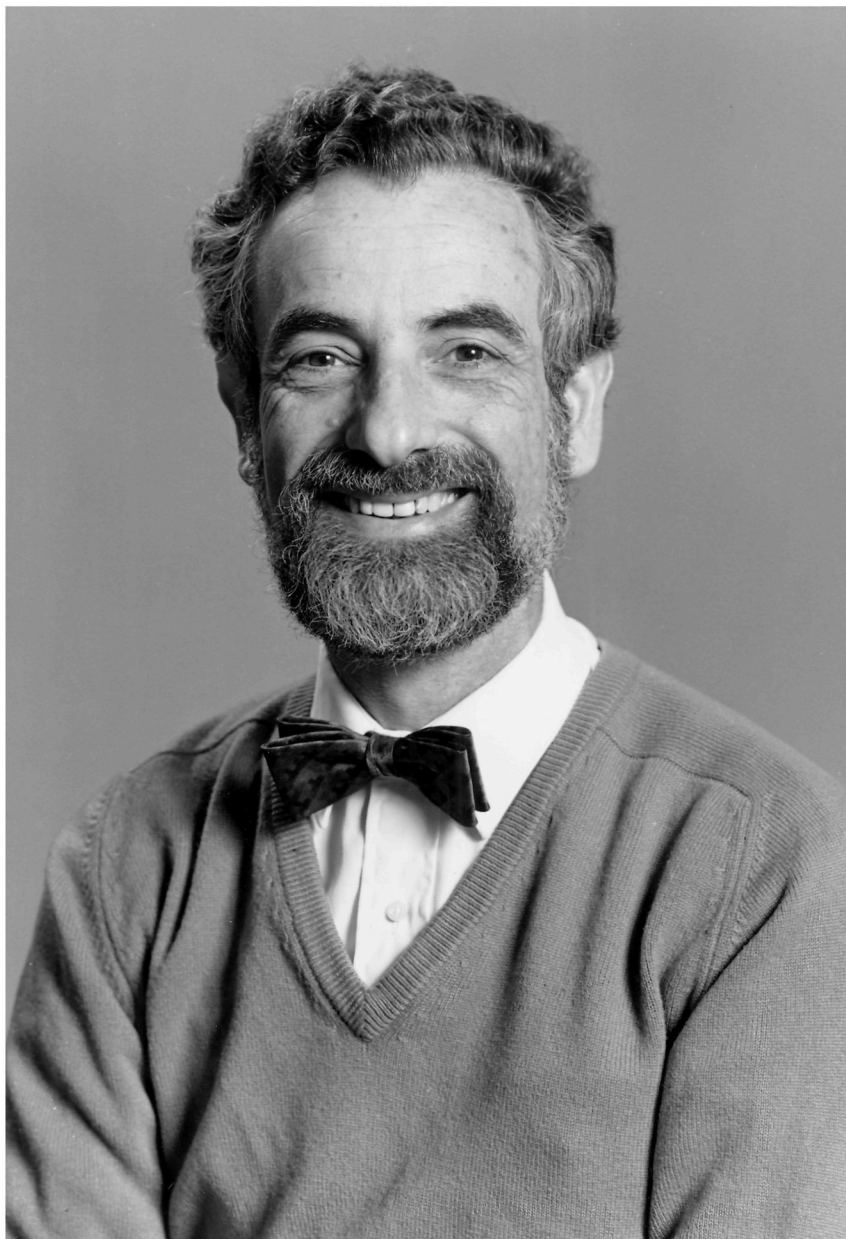
Fourth Edition



Springer

The Stability of Matter:
From Atoms to Stars

Selecta of Elliott H. Lieb



ELLIOTT H. LIEB

The Stability of Matter: From Atoms to Stars

Selecta of Elliott H. Lieb

Edited by W. Thirring

With a Preface by F. Dyson

Fourth Edition



Springer

Professor Elliott H. Lieb
Departments of Mathematics and Physics
Jadwin Hall
Princeton University
P.O. Box 708
Princeton, New Jersey 08544-0708, USA

Professor Walter Thirring
Institut für Theoretische Physik, Universität Wien
Boltzmanngasse 5, 1090 Wien, Austria

Professor Freeman Dyson
The Institute of Advanced Study
School of Natural Sciences
Princeton, New Jersey 08540, USA

Library of Congress Control Number:
2004108033

ISBN 3-540-22212-X Fourth Edition
Springer Berlin Heidelberg New York

ISBN 3-540-42083-5 Third Edition
Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other ways, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

Springer Berlin Heidelberg New York
is a part of Springer science+business media

springeronline.com

© Springer-Verlag Berlin Heidelberg New York 1991, 1997, 2001, 2005
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Printed on acid-free paper SPIN 11013839 55/3141/YL - 5 4 3 2 1 0

Preface to the Fourth Edition

This fourth edition of selecta of my work on the stability of matter contains recent work on two topics that continue to fascinate me: Quantum electrodynamics (QED) and the Bose gas.

Three papers have been added to Part VII on QED. As I mentioned in the preface to the third edition, there must be a way to formulate a *non-perturbative* QED, presumably with an ultraviolet cutoff, that correctly describes low energy physics, i.e., ordinary matter and its interaction with the electromagnetic field. The new paper VII.5, which “quantizes” the results in V.9, shows that the elementary ‘no-pair’ version of relativistic QED (using the Dirac operator) is unstable when many-body effects are taken into account. Stability can be restored, however, if the Dirac operator with the field, instead of the bare Dirac operator, is used to define an electron. Thus, the notion of a “bare” electron without its self-field is physically questionable.

One of the truly basic quantum-mechanical phenomena is the existence of a ground state, but to define this one must have a Hamiltonian and a variational principle. This is not easy to formulate in a truly relativistic theory using the Dirac operator, but work has been done trying to make the ground state clear for the non-relativistic Schrödinger operator coupled to the relativistic electromagnetic field. Paper VII.3 completes the work started in VII.2 by showing that the “binding condition” of VII.2 is satisfied and hence that atoms and molecules have true ground states in this theory, despite the possibility of an “infrared problem”.

Paper VII.4 is a first, primitive attempt to quantify the mass renormalization problem non-perturbatively.

Part VIII on the Bose gas has four new papers. Paper VIII.7 carries forward what was begun in VIII.6 on the one-component gas (‘jellium’) by proving that Dyson’s 1967 conjecture on the ground state of a *two*-component gas of high density charged bosons is exactly describable, asymptotically, by a mean-field equation.

The theory of the low density gas, is quite a different matter and its ground state is shown to have 100% Bose-Einstein condensation (in VIII.8) and 100% superfluidity (in VIII.9) for gases in traps (in the Gross-Pitaevskii limit). These long-conjectured results are now proved, but the ‘holy grail’ of proving Bose-Einstein condensation for an interacting gas in the usual thermodynamic limit remains as a challenge.

Other interesting facts, experimental and theoretical, about dilute gases in traps continue to emerge. Recent experiments show that gases in elongated traps behave like one-dimensional systems, and the mathematical proof is announced in VIII.10. A more detailed discussion of these papers is in VIII.1.

Three other selecta have also been published: *Inequalities*, edited by Michael Loss and Mary Beth Ruskai; *Statistical Mechanics and Condensed Matter Physics*

and Exactly Soluble Models, both edited by Bruno Nachtergaele, Jan Philip Solovej and Jakob Yngvason. Section VIII of the Condensed Matter Physics selecta contains other papers on the Bose gas, including a recent review article.

It is a pleasure to thank Wolf Beiglböck and Walter Thirring for their support and encouragement and Sabine Lehr and Brigitte Reichel-Mayer for their patient help with the production through four editions.

Princeton, August 2004

Elliott Lieb

Preface to the Third Edition

The second edition of this “selecta” of my work on the stability of matter was sold out and this presented an opportunity to add some newer work on the quantum-mechanical many-body problem. In order to do so, and still keep the volume within manageable limits, it was necessary to delete a few papers that appeared in the previous editions. This was done without sacrificing content, however, since the material contained in the deleted papers still appears, in abbreviated form, at least, in other papers reprinted here.

Sections VII and VIII are new. The former is on *quantum electrodynamics* (QED), to which I was led by consideration of stability of the non-relativistic many-body Coulomb problem, as contained in the first and second editions. In particular, the fragility of stability of matter with classical magnetic fields, which requires a bound on the fine-structure constant even in the non-relativistic case (item V.4), leads to the question of stability in a theory with quantized fields. There are many unresolved problems of QED if one attempts to develop a non-perturbative theory – as everyone knows. A non-perturbative theory is essential, however, if one is going to understand the stability of the many-body problem, which is the stability of ordinary matter. Some physicists will say that a non-perturbative QED does not exist – and this might be true in the absence of cutoffs – but an effective theory with cutoffs of a few Mev *must* exist since matter exists.

At the present time physicists believe fully in the non-relativistic Schrödinger equation and even write philosophical tracts about it. This equation is the basis of condensed matter physics and, while it is understood that relativistic effects exist, they are not large and can be treated as mild perturbations. The original Schrödinger picture is alive and well as a very good approximation to reality, and shows no signs of internal inconsistency. For this reason, the ‘stability of matter’ question, which takes the equation literally and seriously without approximation, is accepted as a legitimate question about physical reality.

The radiation field, on the other hand, especially the quantized radiation field, is almost always treated perturbatively in one way or another. The Schrödinger equation, including coupling to the quantized radiation field treated perturbatively, describes ordinary matter with great accuracy, for it is the basis of chemistry and solid state physics.

One can ask if there is a theory, analogous to the Schrödinger theory, that incorporates particles and electromagnetic fields together without having to resort to perturbation theory. If it were really impossible to have a self-consistent, non-perturbative theory of matter and the low frequency radiation field it would not be unreasonable to call this a crisis in physical theory. It has always been assumed that at each level of physical reality there is a self consistent theory that describes

phenomena well at that level. For example, we have chemistry, which is well described at the level of molecules, phase transitions and thermodynamics, which is presumably well described by statistical mechanics, etc. At the present time we have the beginnings of a consistent theory of matter and radiation at the non-relativistic level, but a corresponding relativistic theory, is less developed.

Section VII describe some primitive attempts to address some of the questions that have to be resolved. One problem concerns renormalization. Starting from a relativistic Dirac theory, one finds that the radiation field forces a renormalization of the electron mass. This is well known, but what is not so well known is that the nonrelativistic theory forces an even bigger renormalization. Thus, if one starts from a relativistic theory, performs the renormalization, and then asks for the nonrelativistic limit of such a theory, it appears that one has to make further renormalizations – which it should not be necessary to do.

The new Section VIII is devoted to a subject that was a favorite of mine from the beginning of my career: *many-boson systems*. Some of the papers reprinted in the first and second editions on the stability of charged boson systems have been moved to this section. A very brief summary of all this material, as well as a list of my earliest work on this topic is given in an introduction VIII.1. Three new papers (VIII.2, 3, 4) are about the solution to a very old problem, namely the calculation of the ground state energy of a system of Bose particles with short-range forces at low density. The stability question is trivial, but the calculation of physical quantities in this intensely quantum-mechanical regime requires another sort of physical insight that goes back half a century but which was never rigorously validated.

Currently, the study of low density, low temperature Bose gases in “traps” is an active area of experimental physics. The extension of the homogeneous gas results to inhomogeneous gases in traps is in VIII.3. Similar subtleties arise for charged bosons in a neutralizing background (“jellium”), and this was resolved recently in VIII.7. The new material is summarized in VIII.4.

Certainly, much more remains to be understood about the ground states of bosonic systems. The two-component charged Bose gas, which is related to the jellium problem (see VIII.6) should be better understood, quantitatively. And, of course, there is the ancient quest for a proof of the existence of Bose condensation in interacting systems.

Princeton, March 2001

Elliott Lieb

Preface to the Second Edition

The first edition of “The Stability of Matter: From Atoms to Stars” was sold out after a time unusually short for a selecta collection and we thought it appropriate not just to make a reprinting but to include eight new contributions. They demonstrate that this field is still lively and keeps revealing unexpected features. Of course, we restricted ourselves to developments in which Elliott Lieb participated and thus the heroic struggle in Thomas-Fermi theory where the accuracy has been pushed from $Z^{7/3}$ to $Z^{5/3}$ is not included. A rich landscape opened up after Jakob Yngvason’s observation that atoms in magnetic fields also are described in suitable limits by a Thomas-Fermi-type theory. Together with Elliott Lieb and Jan Philip Solovej it was eventually worked out that one has to distinguish 5 regions. If one takes as a dimensionless measure of the magnetic field strength B the ratio Larmor radius/Bohr radius one can compare it with $N \sim Z$ and for each of the domains

- (i) $B \ll N^{4/3}$,
- (ii) $B \sim N^{4/3}$,
- (iii) $N^{4/3} \ll B \ll N^3$,
- (iv) $B \sim N^3$,
- (v) $B \gg N^3$

a different version of magnetic Thomas-Fermi theory becomes exact in the limit $N \rightarrow \infty$. In two dimensions and a confining potential (“quantum dots”) the situation is somewhat simpler, one has to distinguish only

- (i) $B \ll N$,
- (ii) $B \sim N$,
- (iii) $B \gg N$

and thus there are three semiclassical theories asymptotically exact. These fine distinctions make it clear how careful one has to be when people claim to have derived results valid for high magnetic fields. There is plenty of room for confusion if they pertain to different regions.

The question of stability of matter in a magnetic field B has also been further cleaned up. Already the partial results in V.3 and V.4 of the first edition showed where the problem resides. Whereas diamagnetism poses no problem to stability since there a magnetic field pushes the energy up, the paramagnetism of the electron’s spin magnetic moment can lower the energy arbitrarily much with increasing B . Only if one includes the field energy $+\frac{1}{8\pi} \int B^2$ one gets a bound uniform in B . This has now been shown to be true for arbitrarily many electrons and nuclei provided α and $Z\alpha^2$ are small enough. Though the sharp

constants have not been determined, for $\alpha = 1/137$ and $Z \leq 1050$ (nonrelativistic) stability is guaranteed. Two recent works pertaining to this question have now been included:

1. A transparent and relatively simple proof of stability for Coulomb systems with relativistic kinetic energy is now available.
2. The problem has been studied with the Dirac Hamiltonians for the individual electrons in an external magnetic field. Here the question of how to fill the Dirac sea arises and it turns out that the correct way is to use the Hamiltonian including the magnetic field. Once this is done the Coulomb interaction does not introduce an instability provided the charges are below certain limits. On the contrary if one fills the Dirac sea of the free electrons and then introduces the magnetic field then one gets instability. Amusingly the relativistic theory also allows the above nonrelativistic limit on Z to be improved to $Z \leq 2265$.

Unfortunately QED has not yet matured mathematically to a state where these questions can be answered in a full-fledged relativistic quantum field theory and only in patchworks of relativistic corrections some answers are obtained. Thus this field is by no means exhausted and will keep challenging future generations.

Vienna, November 1996

Walter Thirring

Preface to the First Edition

With this book, Elliott Lieb joins his peers Hermann Weyl and Chen Ning Yang. Weyl's *Selecta* was published in 1956, Yang's *Selected Papers* in 1983. Lieb's "Selecta", like its predecessors, gives us the essence of a great mathematical physicist concentrated into one convenient volume. Weyl, Yang and Lieb have much more in common than the accident of this manner of publication. They have in common a style and a tradition. Each of them is master of a formidable mathematical technique. Each of them uses hard mathematical analysis to reach an understanding of physical laws. Each of them enriches both physics and mathematics by finding new mathematical depths in the description of familiar physical processes.

The central theme of Weyl's work in mathematical physics was the idea of symmetry, linking physical invariance-principles with the mathematics of group-theory. One of Yang's central themes is the idea of a gauge field, linking physical interactions with the mathematics of fibre-bundles. The central theme of Lieb's papers collected in this book is the classical Thomas-Fermi model of an atom, linking the physical stability of matter with the mathematics of functional analysis. In all three cases, a rather simple physical idea provided the starting-point for building a grand and beautiful mathematical structure. Weyl, Yang and Lieb were not content with merely solving a problem. Each of them was concerned with understanding the deep mathematical roots out of which physical phenomena grow.

The historical development of Lieb's thinking is explained in the review articles in this volume, items 65, 92 and 136 in Lieb's publication list. I do not need to add explanatory remarks to Lieb's lucid narrative. I suppose the reason I was asked to write this preface is¹ because Lenard and I found a proof of the stability of matter in 1967. Our proof was so complicated and so unilluminating that it stimulated Lieb and Thirring to find the first decent proof, included here (paper 85 in the publication list). Why was our proof so bad and why was theirs so good? The reason is simple. Lenard and I began with mathematical tricks and hacked our way through a forest of inequalities without any physical understanding. Lieb and Thirring began with physical understanding and went on to find the appropriate mathematical language to make their understanding rigorous. Our proof was a dead end. Theirs was a gateway to the new world of ideas collected in this book.

Princeton, March 1990

Freeman Dyson

Contents

Introduction	1
--------------------	---

Part I. Review

I.1 The Stability of Matter: From Atoms to Stars	11
--	----

Part II. Exact Results on Atoms

II.1 Lower Bound to the Energy of Complex Atoms (with P. Hertel and W. Thirring)	63
II.2 Improved Lower Bound on the Indirect Coulomb Energy (with S. Oxford)	65
II.3 Monotonicity of the Molecular Electronic Energy in the Nuclear Coordinates	79
II.4 Proof of the Stability of Highly Negative Ions in the Absence of the Pauli Principle (with R. Benguria)	83
II.5 Atomic and Molecular Negative Ions	87
II.6 Bound on the Maximum Negative Ionization of Atoms and Molecules	91
II.7 Approximate Neutrality of Large- Z Ions (with I.M. Sigal, B. Simon and W. Thirring)	103
II.8 Universal Nature of van der Waals Forces for Coulomb Systems (with W. Thirring)	113
II.9 Electron Density Near the Nucleus of a Large Atom (with O.J. Heilmann)	121
II.10 Proof of a Conjecture About Atomic and Molecular Cores Related to Scott's Correction (with A. Iantchenko and H. Siedentop)	127
II.11 Asymptotics of Natural and Artificial Atoms in Strong Magnetic Fields (with J.P. Solovej and J. Yngvason)	147
II.12 Ground States of Large Quantum Dots in Magnetic Fields (with J.P. Solovej and J. Yngvason)	171

Part III. General Results with Applications to Atoms

III.1 Kinetic Energy Bounds and Their Application to the Stability of Matter	193
III.2 Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their Relation to Sobolev Inequalities (with W. Thirring)	205

III.3	On Semi-Classical Bounds for Eigenvalues of Schrödinger Operators (with M. Aizenman)	241
III.4	The Number of Bound States of One-Body Schrödinger Operators and the Weyl Problem	245
III.5	Variational Principle for Many-Fermion Systems	257

Part IV. Thomas–Fermi and Related Theories

IV.1	Thomas–Fermi and Related Theories of Atoms and Molecules . . .	263
IV.2	The Hartree–Fock Theory for Coulomb Systems (with B. Simon) .	303
IV.3	There Are No Unfilled Shells in Unrestricted Hartree–Fock Theory (with V. Bach, M. Loss and J.P. Solovej)	313
IV.4	Many-Body Atomic Potentials in Thomas–Fermi Theory (with R. Benguria)	317
IV.5	The Positivity of the Pressure in Thomas–Fermi Theory (with R. Benguria)	329
IV.6	The Thomas–Fermi–von Weizsacker Theory of Atoms and Molecules (with R. Benguria and H. Brezis)	357
IV.7	Analysis of the Thomas–Fermi–von Weizsacker Equation for an Infinite Atom Without Electron Repulsion	371
IV.8	The Most Negative Ion in the Thomas–Fermi–von Weizsacker Theory of Atoms and Molecules (with R. Benguria)	383

Part V. Stability of Matter

V.1	Bound for the Kinetic Energy of Fermions Which Proves the Stability of Matter (with W. Thirring)	401
V.2	Stability of Coulomb Systems with Magnetic Fields. I. The One-Electron Atom (with J. Fröhlich and M. Loss)	405
V.3	Stability of Coulomb Systems with Magnetic Fields. II. The Many-Electron Atom and the One-Electron Molecule (with M. Loss)	425
V.4	Stability of Matter in Magnetic Fields (with M. Loss and J.P. Solovej)	437
V.5	The Chandrasekhar Theory of Stellar Collapse as the Limit of Quantum Mechanics (with H.-T. Yau)	443
V.6	One-Electron Relativistic Molecules with Coulomb Interaction (with I. Daubechies)	471
V.7	The Stability and Instability of Relativistic Matter (with H.-T. Yau)	485
V.8	Stability of Relativistic Matter Via Thomas–Fermi Theory (with M. Loss and H. Siedentop)	523
V.9	Stability and Instability of Relativistic Electrons in Classical Electromagnetic Fields (with H. Siedentop and J.P. Solovej)	535

Part VI. The Thermodynamic Limit for Real Matter with Coulomb Forces

VI.1	The Stability of Matter	561
VI.2	Existence of Thermodynamics for Real Matter with Coulomb Forces (with J.L. Lebowitz)	579
VI.3	The Thermodynamic Limit for Jellium (with H. Narnhofer)	583

Part VII. Quantum Electrodynamics

VII.1	Self Energy of Electrons in Non-perturbative QED (with M. Loss)	607
VII.2	Ground-States in Non-relativistic Quantum Electrodynamics (with M. Griesemer and M. Loss)	625
VII.3	Existence of Atoms and Molecules in Non-Relativistic Quantum Electrodynamics (with M. Loss)	665
VII.4	A Bound on Binding Energies and Mass Renormalization in Models of Quantum Electrodynamics (with M. Loss)	709
VII.5	Stability of a Model of Relativistic Quantum Electrodynamics (with M. Loss)	723

Part VIII. Bosonic Systems

VIII.1	Introduction	753
VIII.2	Ground State Energy of the Low Density Bose Gas (with J. Yngvason)	755
VIII.3	Bosons in a Trap: A Rigorous Derivation of the Gross–Pitaevskii Energy Functional (with R. Seiringer and J. Yngvason)	759
VIII.4	The $N^{5/3}$ Law for Bosons	773
VIII.5	The $N^{7/5}$ Law for Charged Bosons (with J.G. Conlon and H.-T. Yau)	777
VIII.6	Ground State Energy of the One-Component Charged Bose Gas (with J.P. Solovej)	809
VIII.7	Ground State Energy of the Two-Component Charged Bose Gas (with J.P. Solovej)	849
VIII.8	Proof of Bose-Einstein Condensation for Dilute Trapped Gases (with R. Seiringer)	899
VIII.9	Superfluidity in Dilute Trapped Bose Gases (with R. Seiringer and J. Yngvason)	903
VIII.10	One-Dimensional Bosons in Three-Dimensional Traps (with R. Seiringer and J. Yngvason)	909

Selecta of Elliott H. Lieb	913
--------------------------------------	-----

Publications of Elliott H. Lieb	915
---	-----

Introduction

1

“Once we have the fundamental equation (*Urgleichung*¹) we have the theory of everything” is the creed of some physicists. They go on to say that “then physics is complete and we have to seek other employment”. Fortunately, other scientists do not subscribe to this credo, for they believe that it is not the few Greek letters of the *Urgleichung* that are the essential physics but rather that physics consists of all the consequences of the basic laws that have to be unearthed by hard analysis. In fact, sometimes it is the case that physics is not so much determined by the specific form of the fundamental laws but rather by more general mathematical relations. For instance, the KAM-theorem that determines the stability of planetary orbits does not depend on the exact $1/r$ law of the gravitational potential but it has a number theoretic origin. Thus a proper understanding of physics requires following several different roads: One analyzes the general structure of equations and the new concepts emerging from them; one solves simplified models which one hopes render typical features; one tries to prove general theorems which bring some systematics into the gross features of classes of systems and so on. Elliott Lieb has followed these roads and made landmark contributions to all of them. Thus it was a difficult assignment when Professor Beiglböck of Springer-Verlag asked me to prepare selecta on one subject from Lieb’s rich publication list². When I finally chose the papers around the theme “stability of matter” I not only followed my own preference but I also wanted to bring the following points to the fore:

- (a) It is sometimes felt that mathematical physics deals with epsilon-irrelevant to physics. Quite on the contrary, here one sees the dominant features of real matter emerging from deeper mathematical analysis.
- (b) The *Urgleichung* seems to be an ever receding mirage which leaves in its wake laws which describe certain more or less broad classes of phenomena. Perhaps the widest class is that associated with the Schrödinger equation with $1/r$ -potentials, which appears to be relevant from atoms and molecules to bulk matter and even cosmic bodies. Thus the papers reproduced here do not deal with mathematical games but with the very physics necessary for our life.
- (c) In mathematics we see a never ending struggle for predominance between geometry and analysis, the fashions swinging between extremes. Not too long ago the intuitive geometrical way in which most physicists think was scorned by

¹ *Urgleichung* (which is a word coined by Heisenberg) has now been replaced by the TOE (theory of everything) of the advocates of superstring theory.

² Henceforth quoted as [Lieb, ...].

mathematicians and only results from abstract analysis were accepted. Today the pendulum has swung the other way and the admired heroes are people with geometric vision, whereas great analysts like J. von Neumann tend to be thought of as degenerate logicians. As a physicist one should remain neutral with regard to internal affairs of mathematics, but it is always worthwhile to steer against the trend. Stability of matter illustrates beautifully that the great masters have forged for us the very analytical tools which we need to extract the physics from the fundamental equations.

In the selected papers the deeper results are carefully derived and what is presupposed on the reader's part are standard mathematical relations like convexity inequalities. Thus, it is to be hoped that this volume is not only of historical interest but also a useful source for workers in quantum mechanics.

2 Extensivity

We all take for granted that two liters of gasoline contain twice as much energy as one liter. In thermodynamics this fact has been dogmatized by the statement that the energy is an extensive quantity. This means that if $E(N)$ is the energy of the N -particle system then $\lim_{N \rightarrow \infty} E(N)/N$ is supposed to exist, or that the energy per particle approaches a limit in the many-body problem. This is what is meant by "stability of matter" and it was tacitly assumed by people working in this field even though at first sight it seems rather improbable for a system of particles interacting via two-body potentials. The Hamiltonian

$$H_N = \sum_{i=1}^N \frac{p_i^2}{2m_i} + \sum_{i>j} v(x_i - x_j) \quad (1)$$

contains a double sum and one might expect $E(N) \sim N^2$ rather than $E(N) \sim N$. Historically the existence of $\lim_{N \rightarrow \infty} \frac{1}{N} E(N)$ was first seriously studied in the framework of classical statistical mechanics by Van Hove [1] where he had to assume a potential with a hard core. However, real matter, which consists of electrons and nuclei, obeys the laws of quantum mechanics and the dominant interactions are electrostatic or, for cosmic bodies, gravitational. In the latter case the N -dependence of the quantum mechanical ground state energy E_N was first established by Levy-Leblond [2] who proved that in the gravitational case, $v(x_i - x_j) = -\kappa m_i m_j / |x_i - x_j|$, $E_N \sim N^3$ for bosons and $E_N \sim N^{7/3}$ for fermions. Consequently, as was pointed out by Fisher and Ruelle [3], the situation for purely electrostatic interactions, $v(x_i - x_j) = e_i e_j / |x_i - x_j|$, was not so clear. A new chapter in science was opened up when this question was finally settled by Dyson and Lenard [4] who derived in a seminal paper the fundamental result that in this case actually $E_N \sim N$ if all particles of one sign of charge are fermions. If this is not the case and there are bosons of both signs of charge then they could show that E_N is somewhere between $N^{5/3}$ and $N^{7/5}$. It is fairly obvious that there is no chance for stability if the potential is all attractive, but it was rather shocking that electrostatics with its screening property might lead to instability. One might be inclined to consider this as a pathology of the Coulomb potential $1/r$ with its long

range and its singularity and that things would be stable once these troublemakers are removed. That this is not the case is illustrated by the following [5].

Proposition

(2)

For the Hamiltonian

$$H_N = \sum_{i=1}^N \frac{p_i^2}{2m} + \sum_{i>j} e_i e_j v(|x_i - x_j|), \quad e_i = \pm e, \quad \sum_i e_i = 0,$$

the quantum mechanical ground state energy E_N satisfies

- (i) $E_N \sim N$ for fermions, $E_N \sim N^{7/5}$ for bosons if

$$v(r) = \frac{1}{r},$$

- (ii) $E_N \sim N$ for fermions and bosons if

$$v(r) = \frac{1 - e^{-\mu r}}{r},$$

- (iii) $E_N \sim N^2$ for fermions and bosons if

$$v(r) = r e^{-\mu r}.$$

Thus it becomes clear that $E_N \sim N$ happens only under exceptional circumstances and one has to inquire what goes wrong in the case that $E_N \sim N^\gamma$, $\gamma > 1$. It is sometimes said that in this case the thermodynamic limit does not exist and therefore thermodynamics no longer applies. Though there is some truth in this statement, the argument is somewhat superficial and a more detailed analysis is warranted. One might object that in reality one deals with finite systems and the existence of $\lim_{N \rightarrow \infty}$ is only of mathematical interest. This is not quite true because the existence of the limit simply says that for a large but finite system the corresponding quantities are already close to their limiting value and the system can be described by the idealized situation $N \rightarrow \infty$. But for this to be true the kind of limit that is taken is irrelevant. For instance, in the gravitational case one can show the existence of $\lim_{N \rightarrow \infty} N^{-7/3} E_N$ for fermions (or $\lim_{N \rightarrow \infty} N^{-3} E_N$ for bosons). For stars ($N \sim 10^{57}$) one is certainly close to the limit, and their energy is predicted equally well by this kind of $N^{7/3}$ limit as in the stable situation. However I shall now discuss three dominating features which set the case $E_N \sim N$ apart from $E_N \sim N^\gamma$, $\gamma > 1$, and justify the terminology “stability of matter”.

2.1 Relativistic Collapse

The relativistic expression for the kinetic energy $\sqrt{p^2 + m^2}$ weakens the zero point pressure to the extent that the “relativistic” quantum mechanical Hamiltonian

$$H_N^r = \sum_i \sqrt{p_i^2 + m_i^2} + \sum_{i>j} v(x_i - x_j) \quad (3)$$

might cease to be bounded from below if N is sufficiently big. Although this fact was recognized at the beginning of quantum mechanics [6], its dramatic consequence that a star of more than about twice the solar mass will collapse was doubted by leading astrophysicists. Only the experimental discovery of pulsars in the sixties provided overwhelming evidence for this prediction of quantum mechanics and therefore its relevance for cosmic bodies. Thus, there seems to be an essential difference between the relativistic and the nonrelativistic Hamiltonian. Whereas (1) with $1/r$ -potentials is always bounded from below – and instability only implies that its lower bound E_N is not proportional to N – relativistically E_N may be $-\infty$ for all $N > N_c =$ some critical particle number (“relativistic collapse”). Related to this is the fact that nonrelativistically the size of the ground state may shrink with N but it always stays finite, whereas relativistically for $N > N_c$ the system keeps contracting until effects not contained in (3) take over. The two phenomena are related by the following fact which illustrates why instability is catastrophic [7].

Proposition

(4)

If in (1) $v(x)$ scales like $v(\lambda x) = \lambda^{-1}v(x)$ then nonrelativistic instability implies relativistic collapse.

The converse statement is false since nonrelativistically matter is stable for arbitrary values of e , whereas relativistically the two-body Hamiltonian already becomes unbounded from below for $\alpha > \pi/2 = \alpha_c(2)$. Once it was found by Daubechies and Lieb [Lieb, 150] that these stability limits for α become more severe in the many-body situation, the question of stability of Coulomb matter with relativistic kinetic energy appeared more serious because $\lim_{N \rightarrow \infty} \alpha_c(N)$ could conceivably be zero. That this is not so for fermions was first shown by Conlon [8] who proved for nuclear charge $Z = 1$ $\alpha_c(N) > 10^{-200}$ for all N , a value which was subsequently improved by Fefferman and de la Llave [9]. Finally, Lieb and Yau [Lieb, 186] obtained the optimal result which is stability if $\alpha Z \leq 2/\pi$ and $\alpha \leq 1/94$. By Proposition (4) in the other cases (Coulomb matter with positive and negative bosons or with gravitation) the relativistic collapse is unavoidable.

2.2 Thermodynamic Stability

Shortly after $H_N \geq -N \cdot \text{const}$ was established, Lieb and Lebowitz showed the existence of the thermodynamic functions in the limit $N \rightarrow \infty$ for Coulomb matter [Lieb, 43]. This does not follow automatically from stability because the long range of the Coulomb potential poses additional problems. They had to demonstrate that this is sufficiently screened so that separated portions of matter are sufficiently isolated. They not only generalized the work of Van Hove et al. [1,3] to the realistic situation but also considered the microcanonical ensemble and showed that the specific heat was positive. Van Hove had shown that classically the limiting system had positive compressibility which Lieb and Lebowitz also verified for the quantum Coulomb case. These conditions for thermodynamic stability are concisely expressed by a convexity property. Denote by $H_{N,V}$ the N -particle Hamiltonian (1) in a volume V , by $\mathcal{H}_S$ an e^S -dimensional subspace of the Hilbertspace in which $H_{N,V}$ acts, and $\text{Tr}_{\mathcal{H}_S}$ the trace in this subspace. Then for a finite system the energy as function of particle number, entropy and volume is

$$E(N, S, V) = \inf_{\mathcal{H}_S} \text{Tr}_{\mathcal{H}_S} H_{N,V}. \quad (5)$$

In the thermodynamic limit one considers

$$\varepsilon(N, S, V) = \lim_{\lambda \rightarrow \infty} \lambda^{-1} H(\lambda N, \lambda S, \lambda V) \quad (6)$$

and thermodynamic stability requires that ε is jointly convex in its arguments. Lest the reader might think that one gets thermodynamic stability for free one has to note that it was shown by Hertel and Thirring [10] that in the gravitational case the temperature-dependent Thomas-Fermi theory becomes exact in a certain limit such that in this case $\varepsilon(N, S, V)$ not only exists but is calculable. Since this system is not stable, ε has to be defined by

$$\varepsilon(N, S, V) = \lim_{\lambda \rightarrow \infty} \lambda^{-7/3} H(\lambda N, \lambda S, \lambda^{-1} V). \quad (7)$$

It turned out that this system was also thermodynamically unstable; it showed a region of negative specific heat [11,12,13]. Such a phenomenon had been previously discovered in some models [14,15,16] and was always suspected by astrophysicists. That it was no accident that it occurred for the nonextensive Hamiltonian was clarified by Landsberg [17] by means of the following.

Theorem

(8)

Let $x \rightarrow f(x)$ be a map from a convex set of $\mathbf{R}^n$ into $\mathbf{R}$. Then any two of the conditions

$$\begin{aligned} \text{Homogeneity: } & H : f(\lambda x) = \lambda f(x), \quad \lambda \in \mathbf{R}^+ \\ \text{Subadditivity: } & S : f(x_1 + x_2) \leq f(x_1) + f(x_2) \\ \text{Convexity: } & C : f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \end{aligned}$$

imply the third.

If we take for f the function $\varepsilon(N, S, V)$ then H is obviously satisfied in the stable situation (6) but not in the unstable case (7). Thus H is a condition for stability against implosion. On the other hand, the subadditivity S means a stability against explosion; one gains energy by putting two parts together. Finally, C is thermodynamic stability. The theorem says that any two of these stability notions imply the third or, if one fails to hold, the others cannot both be true. In the cases we considered, S holds. This is clear for attractive interactions like gravity whereas for repulsive potentials (e.g. all charges having the same sign) the system would be explosive and S is violated. However, for electrically neutral systems there is always a van der Waals attraction [Lieb, 166] so that S is satisfied, which means that H and C become equivalent. Thus it turns out that our stability condition in terms of extensivity is equivalent to thermodynamic stability provided the system is not explosive.

2.3 The Existence of Quantum Field Theory

Systems with negative specific heat cannot coexist with other systems. They heat up and give off energy until they reach a state of positive specific heat. This is what thermodynamic considerations tell us [15] but it was doubted [18] that in these cases

thermodynamics is applicable and reflects the dynamics of the system. However, recent computer studies [19,20] of the dynamics of unstable systems have revealed that they behave exactly the way one expected by determining the dominant feature in phase space. In these studies one solved the classical equations of motion for 400 particles on a torus with

$$v(x_i - x_j) = -\exp[-|x_i - x_j|^2/b^2]$$

or

$$v(x_i - x_j) = e_i e_j (x_i - x_j)^2 \exp[-|x_i - x_j|^2/b^2].$$

It turned out that irrespective of the initial state, if the total energy is sufficiently small a hot cluster of size b with N_c , say 150 particles, developed, the rest being a homogeneous atmosphere. The temperature increased proportional to N_c . Thus there is no hope for a limiting dynamics for $N \rightarrow \infty$. In particular in quantum field theory, where N is not restricted, one cannot expect that a Hamiltonian H with a smooth potential that, however, leads to an unstable system leads to a dynamics in the Heisenberg representation via the usual formula

$$a(t) = e^{iHt} a e^{-iHt}.$$

Since states with unlimited N will lead to unlimited temperatures and thus unlimited velocities there cannot be a state independent dynamics. These ideas about the relevance of stability for the existence of quantum field theory were brought forward first by Dyson [21] and have been substantiated by proofs only in the converse direction. If one achieves stability by a momentum cut-off then there exists a dynamics in the Heisenberg representation for the quantum field theory [22].

Thus, if one wants a many-body system to behave in the way we are used to, the first question to be answered is its stability. Instability seems to be the rule rather than the exception and actually we owe our lives to the gravitational instability. Not only did Boltzmann's heat death not take place but, on the contrary, the universe, which originally is supposed to have been in an equilibrium state, developed hot clusters, the stars. Thus we can enjoy the sunshine which is rich in energy and lean in entropy, the kind of diet we need.

3

We have seen how a seemingly inconspicuous inequality turned into a key notion for understanding many-body physics. One of the great contributors to this development was Elliott Lieb. He usually did not say the first word on any of these issues but, with various collaborators, the last. His work (with B. Simon) [Lieb, 97] which put Thomas-Fermi (T.F.) theory on a firm basis opened the way for a better understanding of stability of matter. Since they proved that there is no chemical binding in T.F. theory, stability of matter becomes obvious once one knows that the T.F. energies are lower bounds for the corresponding quantum problem. That this is so is still a conjecture but it was shown by E. Lieb (with W. Thirring) [Lieb, 85] that this is true if one changes the T.F.-energies by some (N -independent) numerical factor.

If the electrons were bosons there remained the question whether the lower bound $\sim -N^{5/3}$ or the upper bound $\sim -N^{7/5}$ given by Dyson and Lenard [4] reflects the

true behavior. This question was settled in an unexpected way. Lieb showed that if the masses of the nuclei are ∞ then $-N^{5/3}$ is correct [Lieb, 118] whereas if they are finite he supplied (with J. Conlon and H.-T. Yau) [Lieb, 188] a lower bound $\sim -N^{7/5}$, thereby showing the essential difference between these two cases. This is all the more surprising since in the justification of the Born-Oppenheimer approximation, which says that there is not much difference between big and infinite nuclear mass, usually no reference is made to the fermionic nature of the electrons.

Once one has established the gross features of Coulombic matter reflected by T.F. theory one can worry about finer details not described by it. Of these I will single out two to which Lieb has significantly contributed. One is the question of negative ions (which do not exist in pure T.F. theory but exist if the von Weizsäcker correction to the kinetic energy is added [Lieb, 130]). Ruskai [24] was the first to show that $N(Z)$, the maximal number of electrons bound by a nucleus (in the real quantum theory), is finite. Lieb gave an elegant proof that $N(Z) < 2Z + 1$ [Lieb, 157]. Furthermore he showed (with I. Sigal, B. Simon and W. Thirring) [Lieb, 185] that if the electrons are fermions then $\lim_{Z \rightarrow \infty} N(Z)/Z = 1$. (If the electrons were bosons this limit would be 1.21 [Lieb, 160], [23].) The latter result on real electrons has recently been improved by Fefferman and Seco but the proof of the conjecture $N(Z) < 2 + Z$ is still a challenge for the future.

A typical feature of quantum mechanics which is not included in T.F. theory or any single electron theory (or density functional theory for that matter) is the attraction between neutral objects. There Lieb (with W. Thirring) [Lieb, 166] showed that the Schrödinger equation predicts a potential which is below $-c/(R + R_0)^6$ where R is the distance between their centers and R_0 a length related to their size.

From Proposition (4) it follows that the relativistic Hamiltonian with gravitation becomes unbounded from below if $N > N_c =$ some critical particle number. It remained to be demonstrated that it stays bounded for $N < N_c$ and to determine N_c . Lieb (with W. Thirring) [Lieb, 158] showed that this was actually the case and gave some bounds for N_c which show the expected behavior $\kappa^{-3/2}$ for fermions and κ^{-1} for bosons ($\kappa =$ gravitational constants in units $\hbar = c = m = 1$). Later he sharpened (with H.-T. Yau) [Lieb, 177] these results and showed that, for $\kappa \rightarrow 0$, N_c for fermions is exactly the Chandrasekhar limit.

The papers selected here show how powerful modern functional analysis is in determining the gross features of real matter from the basic quantum mechanical equations. Whereas a century's effort of the greatest mathematicians on the classical N -body problem with $1/r$ potentials only produced results for the simplest special cases, in quantum mechanics the contours of the general picture have emerged with much greater clarity.

The papers are grouped in various subheadings, the first paper being a recent review which will serve as a convenient introduction.

Vienna, November 1990

Walter Thirring

References

- [1] L. Van Hove, *Physica* *15*, 137 (1950).
- [2] J.M. Lèvy-Leblond, *J. Math. Phys.* *10*, 806 (1969).
- [3] M. Fisher, D. Ruelle, *J. Math. Phys.* *7*, 260 (1966).
- [4] F.J. Dyson, A. Lenard, *J. Math. Phys.* *8*, 423 (1967).
- [5] W. Thirring, *A Course in Mathematical Physics*, Vol IV, Springer, Wien, New York, 1980.
- [6] S. Chandrasekhar, *Month. Not. Roy. Ast. Soc.* *91*, 456 (1931).
- [7] W. Thirring, *Naturwissenschaften* *73*, 605 (1986).
- [8] J. Conlon, *Commun. Math. Phys.* *94*, 439 (1984).
- [9] Ch. Fefferman, R. de la Llave, *I. Rev. Math. Iberoamerican* *2*, 119 (1986).
- [10] P. Hertel, W. Thirring, *J. Math. Phys.* *10*, 1123 (1969).
- [11] P. Hertel, W. Thirring, in *Quanten und Felder*, ed. H.P. Dürr, Vieweg, Braunschweig, 1971.
- [12] A. Pflug, *Commun. Math. Phys.* *22*, 83 (1980).
- [13] J. Messer, *J. Math. Phys.* *22*, 2910 (1981).
- [14] D. Lynden-Bell, R. Wood, *Month. Not.* *138*, 495 (1968).
- [15] W. Thirring, *Z. f. Physik* *235*, 339 (1970).
- [16] P. Hertel, W. Thirring, *Ann. of Phys.* *63*, 520 (1971).
- [17] P. Landsberg, *J. Stat. Phys.* *35*, 159 (1984).
- [18] H. Jensen, private communications to W. Thirring.
- [19] A. Compagner, C. Bruin, A. Roelse, *Phys. Rev. A* *39*, 5989 (1989).
- [20] H. Posch, H. Narnhofer, W. Thirring, in *Simulations of Complex Flows*, NATO A.S.I., ed. M. Mareschal (1990) and “Dynamics of Unstable Systems”, UWThPh-1990-2, submitted to *Phys. Rev. A*
- [21] F.J. Dyson, *Phys. Rev.* *85*, 613 (1952).
- [22] H. Narnhofer, W. Thirring, “On Quantum Field Theories with Galilei-Invariant Interactions”, *Phys. Rev. Lett.* *54*, 1863 (1990).
- [23] B. Baumgartner, *J. Phys. A* *17*, 1593 (1984).
- [24] M.B. Ruskai, *Commun. Math. Phys.* *82*, 457 (1982).

Part I

Review

THE STABILITY OF MATTER: FROM ATOMS TO STARS

ELLIOTT H. LIEB

Why is ordinary matter (e.g., atoms, molecules, people, planets, stars) as stable as it is? Why is it the case, if an atom is thought to be a miniature solar system, that bringing very large numbers of atoms together (say 10^{30}) does not produce a violent explosion? Sometimes explosions do occur, as when stars collapse to form supernovae, but normally matter is well behaved. In short, what is the peculiar mechanics of the elementary particles (electrons and nuclei) that constitute ordinary matter so that the material world can have both rich variety and stability?

The law of motion that governs these particles is the quantum (or wave) mechanics discovered by Schrödinger [SE] in 1926 (with precursors by Bohr, Heisenberg, Sommerfeld and others). Everything we can sense in the material world is governed by this theory and some of its consequences are quite dramatic, e.g., lasers, transistors, computer chips, DNA. (DNA may not appear to be very quantum mechanical, but notice that it consists of a very long, thin, complex structure whose overall length scale is huge compared to the only available characteristic length, namely the size of an atom, and yet it is stable.) But we also see the effects of quantum mechanics, without realizing it, in such mundane facts about stability as that a stone is solid and has a volume which is proportional to its mass, and that bringing two stones together produces nothing more exciting than a bigger stone.

The mathematical proof that quantum mechanics gives rise to the observed stability is not easy because of the strong electric forces among the elementary constituents (electrons and nuclei) of matter. The big breakthrough came in the mid sixties when Dyson and Lenard [DL] showed, by a complicated proof, that stability

Received by the editors August 15, 1989.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 81H99, 81M05, 85A15; Secondary 81C99, 82A15.

This paper was presented as the Sixty-Second Josiah Willard Gibbs Lecture on January 11, 1989 at the 95th annual meeting of the American Mathematical Society in Phoenix, Arizona.

Work partially supported by U.S. National Science Foundation grant PHY-85-15288-A03.

is, indeed, a consequence of quantum mechanics. (Part of their motivation came from earlier work by Van Hove, Lee and Yang, van Kampen, Wils, Mazur, van der Linden, Griffiths, Dobrushin, and especially Fisher and Ruelle who formulated the problem and showed how to handle certain well chosen, but unrealistic forces.) This was a milestone but there was room left for improvement since their results had certain drawbacks and did not cover all possible cases; for instance, it turns out that quantum mechanics, which was originally conceived to understand atoms, is also crucial for understanding why stars do not collapse. Another problem was that they proved what is called here stability of the second kind while the existence of the thermodynamic limit (Theorem 3 below), which is also essential for stability, required further work [LL]. The full story has now, two decades later, mostly been sorted out, and that is the subject of this lecture. The answer contains a few surprises, some of which are not even discussed in today's physics textbooks.

No physics background will be assumed of the reader, so Part I reviews some basic facts. Part II contains a synopsis of the aspects of quantum mechanics needed here. Part III treats the simplest system—the hydrogen atom, and Part IV introduces the strange Pauli exclusion principle for many electrons and extends the discussion to large atoms. Part V deals with the basic issue of the stability of matter (without relativistic effects) while Part VI treats hypothetical, but interesting, matter composed of bosons. Part VII treats the problems introduced by the special theory of relativity. Finally, Part VIII applies the results of Part VII to the structure of stars.

PART I. THE PHYSICAL FACTS AND THEIR PREQUANTUM INTERPRETATION

While it is certainly possible to present the whole story in a purely mathematical setting, it is helpful to begin with a brief discussion of the physical situation.

The first elementary constituent of matter to be discovered was the **electron** (J.J. Thomson, 1897). This particle has a **negative electric charge** (denoted by $-e$) and a **mass**, m . It is easy to produce a beam of electrons (e.g., in a television tube) and use it to measure the ratio e/m quite accurately. The measurement of e alone is much trickier (Millikan 1913). The electron can be considered to be a point, i.e., it has no presently discernible geometric structure. Since matter is normally electrically neutral (otherwise we would feel electric fields everywhere), there must also be another

constituent with positive charge. One of the early ideas about this positive object was that it is a positively charged ball of a radius about equal to the radius of an atom, which is approximately 10^{-8} cm. (This atomic radius is known, e.g., by dividing the volume of a solid, which is the most highly compressed form of matter, by the number of atoms in the solid.) The electrons were thought to be stuck in this charged ball like raisins in a cake; such a structure would have the virtue of being quite stable, almost by fiat. This nice picture was destroyed, however, by Rutherford's classic 1903 experiment which showed that the positive entities were also essentially points. (He did this by scattering positively charged helium nuclei through thin metal foils and by showing that the distribution of scattering angles was the same as for the Kepler problem in which the trajectories are hyperbolas; in other words, the scatterers were effectively points—not extended objects.)

The picture that finally emerged was the following. Ordinary matter is composed of two kinds of particles: the point electrons and positively charged **nuclei**. There are many kinds of nuclei, each of which is composed of positively charged **protons** and chargeless **neutrons**. While each nucleus has a positive radius, this radius (about 10^{-13} cm) is so small compared to any length we shall be considering that it can be taken to be zero for our purposes. The simplest nucleus is the single proton (the nucleus of hydrogen) and it has charge $+e$. The number of protons in a nucleus is denoted by z and the values $z = 1, 2, \dots, 92$, except for $z = 43, 61, 85$, are found in nature. Some of these nuclei, e.g., all $84 \leq z \leq 92$, are unstable (i.e., they eventually break apart spontaneously) and we see this instability as naturally occurring radioactivity, e.g., radium. Nuclei with the missing z values 43, 61, 85, as well as those with $92 < z \leq 109$ have all been produced artificially, but they decay more or less quickly [AM]. Thus, the charge of a naturally occurring nucleus can be $+e$ up to $+92e$ (except for 43, 61, 85), but, as mathematicians often do, it is interesting to ask questions about “the asymptotics as $z \rightarrow \infty$ ” of some problems. Moreover, in almost all cases we shall consider here, the physical constraint that z is an integer need not and will not be imposed. The other constituent—the neutrons—will be of no importance to us until we come to stars. They merely add to the mass of the nucleus, for they are electrically neutral. For each given z several possible neutron numbers actually occur in nature; these different nuclei with a common z are called isotopes of each other. For example, when $z = 1$ we have the **hydrogen nucleus** (1 proton) and the deuterium nucleus (1 proton and 1 neutron)

which occur naturally, and the tritium nucleus (1 proton and 2 neutrons) which is artificial and decays spontaneously in about 12 years into a helium nucleus and an electron, but which is important for hydrogen bombs. Isolated neutrons are also not seen naturally, for they decay in about 13 minutes into a proton and an electron.

Finally, the **nuclear mass**, M , has to be mentioned. It satisfies $zM_p \leq M \leq 3zM_p$ where $M_p = 1837m$ is the mass of a proton. Since the nuclear mass is huge compared to the electron mass, m , it can be considered to be infinite for most purposes, i.e., the nuclei can be regarded as fixed points in $\mathbf{R}^3$, although the location of these points will eventually be determined by the requirement that the total energy of the electron-nucleus system is minimized. A similar approximation is usually made when one considers the solar system; to calculate the motion of the planets the sun can be regarded as fixed.

The forces between these constituents of matter (electrons and nuclei) is given by **Coulomb's inverse square law** of electrostatics: If two particles have charges q_1 and q_2 and locations x_1 and x_2 in $\mathbf{R}^3$ then F_1 —the force on the first due to the second—is minus F_2 —the force on the second due to the first—and is given by

$$(1.1) \quad -F_2 = F_1 = q_1 q_2 \frac{(x_1 - x_2)}{|x_1 - x_2|^3}.$$

(Later on, when stars are discussed, the gravitational force will have to be introduced.) If $q_1 q_2 < 0$ then the force is **attractive**; otherwise it is **repulsive**. This force can also be written as minus the gradient (denoted by ∇) of a **potential energy function**

$$(1.2) \quad W(x_1, x_2) = q_1 q_2 \frac{1}{|x_1 - x_2|},$$

that is

$$(1.3) \quad F_1 = -\nabla_1 W \quad \text{and} \quad F_2 = -\nabla_2 W.$$

If there are N electrons located at $\underline{X} = (x_1, \dots, x_N)$ with $x_i \in \mathbf{R}^3$, and k nuclei with positive charges $\underline{Z} = (z_1, \dots, z_k)$ and located at $\underline{R} = (R_1, \dots, R_k)$ with $R_i \in \mathbf{R}^3$, the **total-potential energy function** is then

$$(1.4) \quad W(\underline{X}) = -A(\underline{X}) + B(\underline{X}) + U$$

with

$$(1.5) \quad A(\underline{X}) = e^2 \sum_{i=1}^N V(x_i)$$

$$(1.6) \quad V(x) = \sum_{j=1}^k z_j |x - R_j|^{-1}$$

$$(1.7) \quad B(\underline{X}) = e^2 \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}$$

$$(1.8) \quad U = e^2 \sum_{1 \leq i < j \leq k} z_i z_j |R_i - R_j|^{-1}.$$

The A term is the electron-nucleus attractive potential energy, with $eV(x)$ being the electric potential of the nuclei. B is the electron-electron repulsive energy and U is the repulsive energy of the nuclei. A , B , U and V depend on $\underline{R}$ and $\underline{Z}$, which are fixed and therefore do not appear explicitly in the notation. It is then the case that the force on the i th particle is

$$(1.9) \quad F_i = -\nabla_i W.$$

In the case of an **atom**, $k = 1$ by definition. The case $k > 1$ will be called the **molecular** case, but it includes not only the molecules of the chemist but also solids, which are really only huge molecules.

So far this is just classical electrostatics and we turn next to classical dynamics. Newton's law of motion is (with a dot denoting $\frac{d}{dt}$, where t is the time)

$$(1.10) \quad m\ddot{x}_i = F_i.$$

This law of motion, which is a system of second order differential equations, is equivalent to the following system of first order equations. Introduce the **Hamiltonian function** which is the function on the **phase space** $\mathbf{R}^{6N} = (\mathbf{R}^3 \times \mathbf{R}^3)^N$ given by

$$(1.11) \quad H(\underline{P}, \underline{X}) = \frac{1}{2m} \sum_{i=1}^N p_i^2 + W(\underline{X}).$$

The notation $\underline{P} = (p_1, \dots, p_N)$ with p_i in $\mathbf{R}^3$ is used, and the quantity

$$(1.12) \quad T = \frac{1}{2m} \sum_{i=1}^N p_i^2$$

is called the **kinetic energy**. The equations of motion (1.10) are equivalent to the following first order system in $\mathbf{R}^{6N}$

$$(1.13) \quad \begin{aligned} v_i &\equiv \dot{x}_i = \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial x_i}. \end{aligned}$$

The **velocity** of the i th electron is v_i and p_i is called its **momentum**: $p_i = mv_i$ by the first equation in (1.13).

From (1.13) it will be seen that $H(\underline{P}, \underline{X})$ is constant throughout the motion, i.e., $dH(\underline{P}(t), \underline{X}(t))/dt = 0$. This fixed number is called the **energy** and is denoted by E ; it depends, of course, on the trajectory, and it is important to note that it can take all values in $(-\infty, \infty)$.

Another interesting fact about the flow defined by (1.13), but one which will not be important for us, is that it preserves Lebesgue measure $dx_1 \cdots dx_N dp_1 \cdots dp_N$ on $\mathbf{R}^{6N}$; this is Liouville's theorem and it follows from the fact that the vector field that defines the flow, $(\partial H/\partial p_1, \dots, \partial H/\partial p_N, -\partial H/\partial x_1, \dots, -\partial H/\partial x_N)$, is divergence free. This theorem is one important reason for introducing the Hamiltonian formalism, for it permits a geometric interpretation of classical mechanics and is crucial for ergodic theory and statistical mechanics. The analogue in quantum mechanics turns out to be that quantum mechanical time evolution is given by a one parameter *unitary* group in Hilbert space (see (2.18))—but time evolution will not concern us here.

Consider the simplest possible case, neutral *hydrogen*, with $z = 1$ (a proton) and one electron ($N = 1$ and $k = 1$). With the proton fixed at the origin (i.e., $R_1 = 0$) the Hamiltonian is $p^2/2m - Ze^2|x|^{-1}$ and classical bound orbits (i.e., orbits which do not escape to infinity) of the electron are well known to be the ellipses of Kepler with the origin as a focus. These can pass as close as we please to the proton. Indeed, in the degenerate case the orbit is a radial line segment and in such an orbit the electron passes through the nucleus. One measure of average closeness of the electron to the nucleus in an orbit is the energy E , which is always negative for a bound orbit. Moreover E can be arbitrarily negative because the electron can be arbitrarily close to the nucleus and also have arbitrarily small kinetic energy T . A consequence of this fact is that the hydrogen atom would be physically unstable; in a gas of many atoms another particle or atom could collide with our atom and absorb energy from it. After many such collisions our electron could find itself in a tiny orbit around the nucleus

and our atom would no longer be recognizable as an object whose radius is supposed to be 10^{-8} cm. Each atom would be an infinite source of energy which could be transmitted to other atoms or to radiation of electromagnetic waves.

The problem was nicely summarized by Jeans [J] in his 1915 textbook.

“There would be a very real difficulty in supposing that the (force) law $1/r^2$ held down to zero values of r . For the force between two charges at zero distance would be infinite; we should have charges of opposite sign continually rushing together and, when once together, no force would be adequate to separate them... Thus the matter in the universe would tend to shrink into nothing or to diminish indefinitely in size.”

The inability to account for stable atoms in terms of classical trajectories of pointlike charged particles was the major problem of prequantum physics. Since the existence of atoms and molecules was largely inferential in those days (nowadays we can actually “see” atoms with the tunneling electron microscope), the inability to account for their structure even led some serious people to question their existence—or at least to question the nice pictures drawn by chemists. The main contribution of quantum mechanics was to provide a quantitative theory that “explains” why the electron cannot fall into the nucleus. In brief, when the electron is close to the nucleus its kinetic energy—which could be zero classically—is forced to increase in such a way that the total energy (1.11) goes to $+\infty$ as the average distance $|x|$ goes to zero. This property is known as the **uncertainty principle**.

PART II. QUANTUM MECHANICS IN A NUTSHELL

Schrödinger’s answer to the problem of classical mechanics was the following. While an electron is truly a point particle, its state at any given time cannot be described by a point $x \in \mathbf{R}^3$ and a momentum $p \in \mathbf{R}^3$ (or velocity $v = \frac{1}{m}p$) as in the classical view. Instead *the state of an electron is a (complex valued) function ψ in $L^2(\mathbf{R}^3)$* . Any ψ will do provided it satisfies the normalization condition

$$(2.1) \quad \|\psi\|_2^2 = \int_{\mathbf{R}^3} |\psi(x)|^2 dx = 1.$$

(Actually, this statement is not accurate; an electron has a property called **spin**, and the mathematical expression of this fact is

that ψ is really in $L^2(\mathbf{R}^3; \mathbf{C}^2)$, i.e., $\psi(x)$ is a two-component spinor, $\psi(x) = (\psi_1(x), \psi_2(x))$ with each $\psi_i \in L^2(\mathbf{R}^3)$. This complication—which does not affect the present discussion very much—will frequently be ignored here.) Thus, the state of an electron is a point ψ in an infinite dimensional Hilbert space instead of a point (p, x) in $\mathbf{R}^6$.

The interpretation of ψ , due to M. Born [BM] (Jammer's book [JM] can be consulted for historical details) is that

$$(2.2) \quad \rho_\psi(x) \equiv |\psi(x)|^2$$

is the probability density for finding the electron at x . The expected value of the potential energy $W(x)$ in the state ψ of one electron is then

$$(2.3) \quad W_\psi = \int_{\mathbf{R}^3} \rho_\psi(x) W(x) dx.$$

To localize an electron at a point x_0 would require $|\psi|^2 = \delta_{x_0}$, where δ_{x_0} is Dirac's delta measure, but this is obviously not the square of an $L^2(\mathbf{R}^3)$ function. Thus, it no longer makes sense to speak of an electron unambiguously located at a single point x_0 .

What about the kinetic energy $T = p^2/2m$? To define T we first consider the Fourier transform of ψ scaled by the number h and defined by

$$(2.4) \quad \hat{\psi}(p) = h^{-3/2} \int_{\mathbf{R}^3} \psi(x) \exp\left[-\frac{2\pi i}{h}(p, x)\right] dx.$$

Usually h is taken to be 1 or 2π in textbooks but any h can be used. Although $\hat{\psi}$ depends on the choice of h , Plancherel's formula is always the following:

$$(2.5) \quad \int_{\mathbf{R}^3} |\hat{\psi}(p)|^2 dp = 1.$$

The constant h relevant for quantum mechanics is Planck's constant; it is not arbitrary and has to be determined experimentally in order to satisfy the following physical interpretation of $|\hat{\psi}(p)|^2$. In view of (2.5), $|\hat{\psi}(p)|^2$ can be regarded as a probability density – it is the probability density of the electron having a momentum p . With this interpretation, the kinetic energy of the electron in the state ψ is

$$(2.6) \quad T_\psi = \frac{1}{2m} \int_{\mathbf{R}^3} |\hat{\psi}(p)|^2 p^2 dp = \frac{1}{2m} \hbar^2 \int_{\mathbf{R}^3} |\nabla \psi(x)|^2 dx,$$

where, in a customary notation, $\hbar = h/2\pi$. The role of h in the definition (2.4) is brought into focus by the right side of (2.6). Numerically $\hbar = 1.05 \times 10^{-27}$ erg seconds.

Another way to formulate (2.3) to (2.6) is to think of $L^2(\mathbf{R}^3)$ as a Hilbert space with inner product $(\psi, \phi) = \int \overline{\psi}(x)\phi(x) dx$ so that

$$(2.7) \quad W_\psi = (\psi, W\psi).$$

The momentum p is replaced by the self adjoint operator

$$p = -i\hbar\nabla \quad \text{and} \quad p^2 = -\hbar^2\Delta$$

with Δ being the Laplacian $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. Then the kinetic energy is (after integrating by parts)

$$(2.8) \quad 2mT_\psi = (p\psi, p\psi) = \hbar^2 \|\nabla\psi\|_2^2 = (\psi, p^2\psi)$$

These concepts can be generalized to the **many-body case** of N electrons by replacing $\mathbf{R}^3$ by $\mathbf{R}^{3N}$. Then $\psi(x)$ is replaced by $\psi(\underline{X}) = \psi(x_1, \dots, x_N)$ in $\bigotimes_1^N L^2(\mathbf{R}^3) \approx L^2(\mathbf{R}^{3N})$. (If spin is included then we use $\bigotimes_1^N L^2(\mathbf{R}^3, \mathbf{C}^2) \approx L^2(\mathbf{R}^{3N}; \mathbf{C}^{2^N})$.) A different generalization that might leap to the reader's mind would be to replace $\psi(x)$ by an N -tuple of functions on $\mathbf{R}^3$; that would lead to a nonlinear theory and would be wrong. The normalization condition is

$$(2.9) \quad \int_{\mathbf{R}^{3N}} |\psi(\underline{X})| dx_1 \cdots dx_N = 1$$

and $|\psi(\underline{X})|^2$ is the joint probability density for finding the electrons at $x_1, \dots, x_N$. The potential and kinetic energies are given by

$$(2.10) \quad W_\psi = \int_{\mathbf{R}^{3N}} W(\underline{X}) |\psi(\underline{X})|^2 dx_1 \cdots dx_N$$

$$(2.11) \quad T_\psi = \frac{1}{2m} \hbar^2 \sum_{i=1}^N \int_{\mathbf{R}^{3N}} |(\nabla_{x_i} \psi)(\underline{X})|^2 dx_1 \cdots dx_N.$$

The **energy** of ψ is then

$$(2.12) \quad E_\psi = T_\psi + W_\psi.$$

The problem that will concern us here is to find a lower bound to the **ground state energy**

$$(2.13) \quad E = E(N, k, \underline{Z}, \underline{R}) = \inf_{\psi} E_\psi,$$

because the boundedness of E precludes the kind of collapse that Jeans was worried about. (Recall that E_ψ depends on $N, k, \underline{Z}$ and $\underline{R}$ since $W(\underline{X})$ does.) In (2.13) ψ is required to satisfy the normalization condition (2.9). It will also be required to satisfy another condition—the **Pauli exclusion principle**—which will be explained later.

The **absolute ground state energy** is

$$(2.14) \quad E(N, k, \underline{Z}) = \inf_{\underline{R}} E(N, k, \underline{Z}, \underline{R})$$

and it is the ground state energy when the nuclei are placed in the most favorable locations.

There are two notions of stability

(A) **STABILITY OF THE FIRST KIND.** $E(N, k, \underline{Z}, \underline{R})$ is finite for every $N, k, \underline{Z}$ and $\underline{R}$.

(B) **STABILITY OF THE SECOND KIND.** There is a nonnegative function $z \mapsto A(z)$ such that for all N and k

$$(2.15) \quad E(N, k, \underline{Z}) \geq -A(z)(N + k)$$

provided each $z_j \leq z$. Recall that $z_j \leq 92$ in nature. The significance of the linear law (2.15) will be discussed later in Part V.

The infimum in (2.13) may or may not be attained by some ψ . If it is, a minimizing ψ , which may not be unique, is called a **ground state**. It satisfies the linear Euler-Lagrange equation (also known as the **time independent Schrödinger equation**)

$$(2.16) \quad H\psi = E\psi$$

where H is the **Schrödinger Hamiltonian**, namely the selfadjoint second order elliptic differential operator

$$(2.17) \quad H = -\frac{1}{2m}\hbar^2 \sum_{i=1}^N \Delta_{x_i} + W(\underline{X}).$$

In the previous Hilbert space notation, $E_\psi = (\psi, H\psi)$.

Quantum mechanics is the study of the eigenvalues of H , i.e., solutions of $H\psi = \lambda\psi$ with ψ in L^2 , and also the associated (time dependent) Schrödinger equation of the time evolution

$$(2.18) \quad H\psi = -i\hbar \frac{\partial \psi}{\partial t}$$

which is important in several contexts, especially in the subject of scattering theory. The eigenvalues λ are called **energy levels** and differences between any two, $\lambda_i - \lambda_j$, is the energy carried away or absorbed by a photon when the system makes a “quantum jump”

from eigenstate ψ_1 to eigenstate ψ_2 . The fact that the system “jumps” from one eigenstate to another is a mysterious axiom of quantum mechanics (more precisely, quantum measurement theory) that need not concern us here, but it has the following important physical implication. According to Einstein, the frequency ν of a photon of energy e is always $\nu = e/h$. Thus, by energy conservation the photon emitted or absorbed has frequency $\nu = (\lambda_i - \lambda_j)/h$. Since the λ_i 's are discrete, the frequencies (or spectral lines) are also discrete. With classical mechanics we would expect to find *all* frequencies, and the experimental existence of discrete spectral lines is regarded in many textbooks as being the crucial problem that classical physics could not explain. However, Jeans's question about the existence of stable atoms capable of emitting these discrete spectral lines can evidently be regarded as a question of higher priority.

In this lecture we shall be interested only in the lowest eigenvalue of H , which is E . Equation (2.18) is therefore irrelevant here. Also (2.16) will not be invoked because we shall deal directly with the minimization problem (2.13) which, incidentally, avoids the technically difficult question of the selfadjointness of H .

PART III. THE HYDROGENIC ATOM AND SOBOLEV'S INEQUALITY

In Parts III-VI we shall use units in which $\hbar^2/2m = \hbar^2/8\pi^2m = 1$ and $e = 1$. This can always be done by changing the length scale $x \rightarrow (\hbar^2/2me^2)x'$. The original energy E' is related to the new energy E' (which will henceforth be denoted simply as E) by $E = (2me^2/\hbar^2)E'$.

The problem for hydrogen is to minimize

$$(3.1) \quad E_\psi = T_\psi + W_\psi = \int_{\mathbf{R}^3} |\nabla \psi(x)|^2 dx - z \int_{\mathbf{R}^3} |\psi(x)|^2 |x|^{-1} dx$$

with $\int |\psi|^2 = 1$. We can take ψ to be real, as is true in all the problems considered here, because the real and imaginary parts of ψ appear independently in (3.1) and (2.1). Of course $z = 1$ for hydrogen but it will be useful later to consider (3.1) with arbitrary $z > 0$.

The fact that T_ψ controls W_ψ , thereby making E finite, is most clearly seen with the Sobolev inequality

$$(3.2) \quad T_\psi = \|\nabla \psi\|_2^2 \geq S \|\psi\|_6^2 = S \left\{ \int_{\mathbf{R}^3} \rho_\psi(x)^3 dx \right\}^{1/3} = S \|\rho_\psi\|_3$$

for any $\psi \in L^2(\mathbf{R}^3)$ (not necessarily normalized). Here $S = 3(\pi/2)^{4/3} \approx 5.5$ and we recall (2.2). The important point about (3.2) is that it bounds T_ψ in terms of ρ_ψ ; the derivatives have been eliminated. The ground state energy can be bounded from below using (3.2), and only the fact that $\int \rho_\psi = 1$, by

$$(3.3) \quad E \geq \inf_{\rho} \left\{ S \|\rho\|_3 - z \int_{\mathbf{R}^3} |x|^{-1} \rho(x) dx \right\}$$

where the infimum is over all nonnegative ρ with $\int \rho = 1$. This minimization problem is an easy exercise and the answer is

$$(3.4) \quad E \geq -\frac{1}{3} z^2.$$

This is a remarkably good bound since the correct answer, which is obtained by solving (2.16), is $E = -z^2/4$.

Before moving on, it is important to understand heuristically why (3.1) leads to $E \approx -z^2$. If ψ is some nice function with $\|\psi\|_2 = 1$ and whose width is about L (for example, the Gaussian $\psi(x) = (\text{const.}) \exp[-x^2/L^2]$), then T_ψ is roughly L^{-2} . On the other hand, W_ψ is roughly $-zL^{-1}$. Therefore, the problem of minimizing E_ψ is roughly the same as

$$(3.5) \quad E \approx \min_L \left\{ \frac{1}{L^2} - \frac{z}{L} \right\}$$

The minimizing L is $L = 2/z$ and the minimum in (3.5) is $-z^2/4$, which is the right answer. Indeed, the minimizing ψ for (3.1) is

$$(3.6) \quad \phi(x) = (\text{const.}) e^{-z|x|/2}$$

which shows that the width of ψ is, in fact, about z^{-1} . From this calculation, and recalling that $|\psi(x)|^2$ is the probability density for the electron, we learn that *the size of a one-electron atom with nuclear charge z decreases with z like z^{-1}* .

There is another version of (3.2) that is technically weaker, but more transparent and eventually more useful. Since $\int \rho_\psi = 1$, a simple application of Hölder's inequality to the right side of (3.2) yields

$$(3.7) \quad T_\psi \geq K^1 \int_{\mathbf{R}^3} \rho_\psi(x)^{5/3} dx.$$

with $K^1 = S$. The constant K^1 can be improved from S to 9.57, but that is conceptually unimportant. Instead of (3.3) we then

have

$$(3.8) \quad E \geq \inf_{\rho} \left\{ \int_{\mathbf{R}^3} [K^1 \rho(x)^{5/3} - |x|^{-1} \rho(x)] dx \right\}$$

and this yields a result only slightly worse than (3.4), namely $-0.35z^2$ with $K^1 = 9.57$. The important point is that (3.7) (with K^1 or with S) is just as good for our purposes as (3.3) because it yields the z^2 law, but the important advantage of (3.7) over (3.3) is that the integral comes with the exponent 1 in (3.7) instead of the exponent $1/3$ in (3.2). This gives (3.7) the form of the integral of an energy density $\rho_{\psi}(x)^{5/3}$. Independent “bumps” in ρ_{ψ} give additive contributions to the right side of (3.7); in (3.2), by contrast, we have to take the cube root—which spoils the additivity.

The inequality (3.7) can be stated poetically as follows. *An electron is like a rubber ball, or a fluid, with an energy density proportional to $\rho_{\psi}^{5/3}$. It costs energy to squeeze it and this accounts for the stability of atoms.*

PART IV. LARGE ATOMS AND THE PAULI EXCLUSION PRINCIPLE

Suppose that $N = z$ and z is large, say $z = 50$. Since $N > 1$, W_{ψ} has two terms which come from $-A$ and B in (1.4). The constant U is zero since there is only one nucleus. The B term is positive but it turns out to be small compared to the A term and, for a preliminary orientation, can be neglected. We would then try to minimize

$$(4.1) \quad \tilde{E}_{\psi} = \sum_{i=1}^N \int_{\mathbf{R}^{3N}} \{ |\nabla_{x_i} \psi(\underline{X})|^2 - z|x_i|^{-1} |\psi(\underline{X})|^2 \} dx_1 \cdots dx_N$$

The minimum with $\|\psi\|_2 = 1$ is then achieved by

$$(4.2) \quad \psi(\underline{X}) = \prod_{i=1}^N \phi(x_i),$$

where $\phi(x)$ is the minimizer for (3.1) and is given in (3.6). This leads to

$$(4.3) \quad \tilde{E} = \inf_{\psi} \tilde{E}_{\psi} = -Nz^2/4 = -z^3/4.$$

From this calculation we would be led to two conclusions: (1) The energy of an atom is finite (stability of the first kind) but it grows like z^3 . (ii) Because the scale of ϕ is z^{-1} , the size of a

large atom decreases with z like z^{-1} . Conclusion (ii) is especially troublesome: *large atoms are smaller than small atoms*.

Both conclusions are false because of an additional axiom introduced by Pauli [P] in 1925 (and interpreted in the context of Schrödinger's wave mechanics by Dirac and Heisenberg in (1926)), namely the **exclusion principle**. This states that the allowed ψ 's in the minimization problem (2.13) are the **antisymmetric functions** in $L^2(\mathbf{R}^{3N}) \approx \bigotimes^N L^2(\mathbf{R}^3)$. (The reason for the word exclusion will be explained shortly.) In other words $\psi \in \bigwedge^N L^2(\mathbf{R}^3)$, the antisymmetric tensor product of $L^2(\mathbf{R}^3)$. More explicitly

$$\begin{aligned} \psi(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = \\ - \psi(x_1, \dots, x_j, \dots, x_i, \dots, x_N) \end{aligned}$$

for every $i \neq j$. If spin is taken into account then the condition is $\psi \in \bigwedge^N L^2(\mathbf{R}^3; \mathbf{C}^2)$. (Equivalently, ψ is an antisymmetric function in $L^2(\mathbf{R}^3; \mathbf{C}^{2^N})$.) The physical significance of this peculiar antisymmetry restriction will not be discussed here, but it is related to the fact that electrons are indistinguishable.

A function in $\bigwedge^N L^2(\mathbf{R}^3; \mathbf{C}^2)$ may be hard to think about—in which case the reader should ignore the complication of spin and think instead about $\bigwedge^N L^2(\mathbf{R}^3)$ without any essential loss. The following remarks may be useful, however. Elementary manipulations with the symmetric (or permutation) group S_N show that our minimization problem (2.13) (or even the Schrödinger equation (2.16) or (2.18)) in $\bigwedge^N L^2(\mathbf{R}^3; \mathbf{C}^2)$ is equivalent to solving the problem in a subset of the more familiar space $L^2(\mathbf{R}^{3N})$. This is the subset consisting all function $\psi = \psi(x_1, \dots, x_N)$ with the property that there is some integer $1 \leq J \leq N$ such that ψ is antisymmetric in the variables $x_1, \dots, x_J$ and also antisymmetric in the variables $x_{J+1}, \dots, x_N$ if $J < N$. No assertion is made about permuting variables x_i and x_j when $i \leq J < j$. Thus, it is as though there are two species of *spinless* particles (J of one kind and $N - J$ of the other kind), each of which satisfies the Pauli exclusion principle for spinless particles. The terminology that is employed to describe these “two kinds” of electrons is “electrons with spin up” and “electrons with spin down”.

Particles such as electrons, protons and neutrons (and some, but not all nuclei) that satisfy the Pauli exclusion principle are called **fermions**. They always have spin. There is another kind of particle in nature—**bosons**. These have the restriction that ψ is symmetric (there is only $L^2(\mathbf{R}^3)$ this time, not $L^2(\mathbf{R}^3; \mathbf{C}^2)$). Every elementary particle is *either* a boson or a fermion. Fortunately, there are

no negatively charged bosons which do not decay rapidly, for matter composed of negative bosons and positive bosons is not stable of the second kind—as will be seen later in Part VI.

It is a very general fact that the infimum in (2.13) over the full tensor product $\otimes^N L^2(\mathbf{R}^3)$ is the infimum over symmetric ψ 's, i.e., bosons. To prove this fact, first replace ψ (which is real and normalized) by $|\psi|$, which does not increase E_ψ because $|\psi|^2$ and $|\nabla\psi|^2$ remain the same. Next, for *any* function $\Phi(X)$ we can construct a symmetric function $\Phi_\sigma(X)$ by summing over permutations, i.e.,

$$\Phi_\sigma(X) = (N!)^{-1} \sum_{\pi \in S_N} \Phi(\pi X).$$

One easily checks that Φ_σ is automatically orthogonal in $\otimes^N L^2(\mathbf{R}^3)$ to the function $\Phi_\alpha \equiv \Phi - \Phi_\sigma$. For the same reason (symmetrizing the integrand in (2.10), (2.11)) E_Φ splits as $E_\Phi = E_{\Phi_\alpha} + E_{\Phi_\sigma}$. Applying this construction to our ψ we easily conclude that

$$\begin{aligned} \|\psi_\sigma\|_2^2 &= (N!)^{-2} \sum_{\mu, \pi \in S_N} (|\psi|(\mu), |\psi|(\pi)) \\ &\geq (N!)^{-2} \sum_{\pi \in S_N} \|\psi\|_2^2 = (N!)^{-1} \|\psi\|_2^2 = (N!)^{-1}. \end{aligned}$$

Let E be the unrestricted infimum and let E^σ denote the infimum restricted to symmetric functions. Then $E^\sigma - E \geq 0$ and

$$\begin{aligned} (4.4) \quad E &= \inf_{\psi} E_\psi = \inf_{\psi} E_{|\psi|} = \inf_{\psi} \{E_{|\psi|_\alpha} + E_{|\psi|_\sigma}\} \\ &\geq \inf_{\psi} \{\|\psi|_\alpha\|_2^2 E + \|\psi|_\sigma\|_2^2 E^\sigma\} \\ &= \inf_{\psi} \{(1 - \|\psi|_\sigma\|_2^2) E + \|\psi|_\sigma\|_2^2 E^\sigma\} \geq E + (N!)^{-1} [E^\sigma - E]. \end{aligned}$$

Hence $E = E^\sigma$ as claimed.

In brief, *the imposition of the Pauli exclusion principle raises E* . The miracle is that it raises E enough so that stability of the second kind holds. While it is easy to state that ψ must be antisymmetric (here we return to $\wedge^N L^2(\mathbf{R}^3)$ for simplicity) it is not easy to quantify the effect of antisymmetry. Even the experts have difficulty, for it is not easy to think of an antisymmetric function of a large number of variables.

The most dramatic effect of antisymmetry concerns T_ψ . In Part III we used the Sobolev inequality to bound T_ψ in terms of

$\rho_\psi(x) = |\psi(x)|^2$ and we would like now to find a similar bound in the N -particle case. It is *not* useful to define $\rho_\psi(\underline{X}) = |\psi(\underline{X})|^2$ because this is a function of N variables and is quite unmanageable. A more useful definition is the **one-particle density** $\rho_\psi : \mathbf{R}^3 \rightarrow \mathbf{R}^+$ given by

$$(4.5) \quad \rho_\psi(x) = N \int_{\mathbf{R}^{3(N-1)}} |\psi(x, x_2, \dots, x_N)|^2 dx_2 \dots dx_N,$$

from which it follows that $\int_{\mathbf{R}^3} \rho_\psi = N$ and $\rho_\psi(x)$ is the **density of electrons** at $x \in \mathbf{R}^3$. (Note that, since ψ is antisymmetric, $|\psi|^2$ is symmetric and therefore it is immaterial which variable is set equal to x in (4.5).)

It is an easy exercise using Minkowski's inequality to deduce from (3.2) that

$$(4.6) \quad T_\psi \geq S \|\rho_\psi\|_3$$

in the N -particle case, but this inequality is not very useful; it does not distinguish the fact that ψ is antisymmetric instead of symmetric. The right side of (4.6) has the following property. Suppose ρ is a smooth nonnegative function on $\mathbf{R}^3$ with $\int \rho = 1$ and suppose also that ρ_ψ for an N particle ψ is $N\rho$. Then the right side of (4.6) is proportional to N . Without the imposition of antisymmetry nothing more can be asserted, but with it T_ψ grows like $N^{5/3}$, as Theorem 1 below shows.

Let us return to (3.7), the weakened form of Sobolev's inequality for one variable. Again, Minkowski's inequality can be used to translate (3.7) into an inequality for N variables:

$$(4.7) \quad T_\psi \geq N^{-2/3} K^1 \int_{\mathbf{R}^3} \rho_\psi(x)^{5/3} dx.$$

The right side of (4.7) also shows a linear dependence on N when $\rho_\psi = N\rho$. But antisymmetry comes to the rescue in the form of the following inequality of Lieb and Thirring [LT1, LT2, L1, L2].

Theorem 1 (The 5/3 law for the kinetic energy of fermions). *There is a universal constant K such that for all N and all antisymmetric (complex) $\psi \in L^2(\mathbf{R}^{3N})$ with $\|\psi\|_2 = 1$*

$$(4.8) \quad T_\psi \geq 2^{2/3} K \int_{\mathbf{R}^3} \rho_\psi(x)^{5/3} dx.$$

If ψ is an antisymmetric function in $L^2(\mathbf{R}^{3N}; \mathbf{C}^{2^N})$ then

$$(4.9) \quad T_\psi \geq K \int_{\mathbf{R}^3} \rho_\psi(x)^{5/3} dx.$$

$K = (2.7709)2^{-2/3} = 1.7455$ will do.

Contrast (4.9) with (4.7); the $N^{-2/3}$ is gone. The poetic remark at the end of Part III is still correct for many fermions, but ρ_ψ is now the total density which, very heuristically, is N times the density of one particle.

Remarks. (1) The sharp K is not known; it is conjectured [LT2] to be $K^c = 3(3\pi^2)^{2/3}/5 = 5.7425$. This number is the “classical” value of K and it arises in the following way. Take a cube $\Gamma_L \subset \mathbf{R}^3$ of length L and then compute $\tau(n, L)$, the minimum of T_ψ over all antisymmetric n -particle ψ ’s with support in $(\Gamma_L)^n$. The best ψ has the form (4.10) below in which each ϕ_i satisfies $-\Delta\phi_i = \lambda_i\phi_i$ in Γ_L and $\phi_i = 0$ on the boundary of Γ_L . One finds, for large n , that $\tau(n, L) \approx 2^{2/3}K^c$ (resp. K^c) times $n^{5/3}/L^2$ and $\rho_\psi(x) \approx n/L^3$ for $x \in \Gamma_L$ and $\rho_\psi(x) = 0$ for $x \notin \Gamma_L$. These values of $T_\psi = \tau(n, L)$, ρ_ψ and K^c then give equality in (4.8) and (4.9).

(2) Theorem 1 is stated above for the case in which n (the dimension of each variable) is 3. A similar theorem holds for all $n \geq 1$ (unlike Sobolev’s inequality which holds only when $n \geq 3$) if $5/3$ is replaced by $1 + 2/n$ and if K is replaced by a constant K_n depending on n (but not on N). See [LT2 and L2].

To explore the significance of Theorem 1, and also the significance of the word “exclusion”, let us examine the simplest kind of antisymmetric ψ , namely a **determinantal function**. Let $\phi_1, \dots, \phi_N$ be any set of N orthonormal functions in $L^2(\mathbf{R}^3)$ and let

$$(4.10) \quad \psi(\underline{X}) = (N!)^{-1/2} \det\{\phi_i(x_j)\}_{i,j=1}^N.$$

This ψ is antisymmetric and has $\|\psi\|_2 = 1$. The word “exclusion” comes from the fact that the ϕ_i ’s have to be distinct—in fact orthogonal to each other. The right side of (4.10) is the closest we can come to thinking in terms of “one $L^2(\mathbf{R}^3)$ function for each electron”. When a physicist or chemist uses a determinantal ψ in calculations the ϕ_i ’s are called *orbitals*, and one says that “only one electron can be in each orbital”. (In the case of $L^2(\mathbf{R}^3; \mathbf{C}^2)$ one says that it is possible to have “at most two electrons in each orbital”.) In fact it was on the basis of this simple heuristic applied to Bohr’s “old quantum mechanics” that Pauli [P] was able to explain the periodic table of the elements in 1925—before Schrödinger invented his equation and hence before the relation between antisymmetry and “exclusion” was understood.

For the ψ given by (4.10), the ρ_ψ in (4.5) and T_ψ are easily computed to be

$$(4.11) \quad \begin{aligned} \rho_\psi(x) &= \sum_{i=1}^N |\phi_i(x)|^2 = \sum_{i=1}^N \rho_{\phi_i}(x) \\ T_\psi &= \sum_{i=1}^N \int_{\mathbf{R}^3} |\nabla \phi_i(x)|^2 dx = \sum_{i=1}^N T_{\phi_i}. \end{aligned}$$

For each ϕ_i , (3.7) says that

$$(4.12) \quad \int_{\mathbf{R}^3} |\nabla \phi_i(x)|^2 dx \geq K^1 \int_{\mathbf{R}^3} |\phi_i(x)|^{10/3} dx.$$

As an illustration of Theorem 1, consider the case that $|\phi_i(x)|^2 = \rho(x)$ for each i , where ρ is a given nonnegative function with $\int \rho = 1$. We could take $\phi_1(x) = \rho(x)^{1/2}$, but then the remaining ϕ_j 's would have to be $\phi_j(x) = \rho(x)^{1/2} \exp[i\theta_j(x)]$, with the θ_j 's real and chosen to insure orthogonality. As N increases, the θ_j 's have to "wiggle" more and more to insure this orthogonality, so that T_{ϕ_j} increases with j . How fast? According to Theorem 1, with $\rho_\psi = N\rho$ in this case,

$$(4.13) \quad T_\psi \geq KN^{5/3} \int_{\mathbf{R}^3} \rho(x)^{5/3} dx$$

and this shows that T_{ϕ_j} increases with j like $j^{2/3}$, on the average.

Armed with Theorem 1, let us return to the question of finding a lower bound to E ; this time the B term will not be ignored. Using (4.9) we have for any admissible ψ ,

$$(4.14) \quad E_\psi \geq K \int_{\mathbf{R}^3} \rho_\psi(x)^{5/3} dx + \int W(\underline{X}) |\psi(\underline{X})|^2 dx_1 \cdots dx_N$$

$$(4.15) \quad \begin{aligned} &= K \int_{\mathbf{R}^3} \rho_\psi(x)^{5/3} dx - z \int_{\mathbf{R}^3} |x|^{-1} \rho_\psi(x) dx \\ &\quad + \int B(\underline{X}) |\psi(\underline{X})|^2 dx_1 \cdots dx_N. \end{aligned}$$

Note that the A term of (1.4) is simply expressible in terms of ρ_ψ , as in (4.15). The B term is more complicated. One would guess by analogy with a fluid of density ρ_ψ that roughly

$$(4.16) \quad \int B(\underline{X}) |\psi(\underline{X})|^2 dx_1 \cdots dx_N = \frac{1}{2} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho_\psi(x) \rho_\psi(y) |x - y|^{-1} dx dy.$$

This is not exactly correct, of course, but it can be proved [LO] that for every normalized ψ (with or without the Pauli exclusion principle), the left side minus the right side of (4.16) is bounded below by $-(1.68) \int \rho_\psi^{4/3}(x) dx$. This error term is small; in any case it can be bounded using the Schwarz inequality by $N^{1/2} (\int \rho_\psi^{5/3})^{1/2}$ and this can easily be controlled by the kinetic energy term $\int \rho_\psi^{5/3}$. Therefore, up to controllable errors

$$E_\psi \geq \mathcal{E}^{TF}(\rho_\psi),$$

where

$$(4.17) \quad \begin{aligned} \mathcal{E}^{TF}(\rho) = & K \int_{\mathbf{R}^3} \rho(x)^{5/3} dx - z \int_{\mathbf{R}^3} |x|^{-1} \rho(x) dx \\ & + \frac{1}{2} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x - y|^{-1} dx dy. \end{aligned}$$

This functional, $\mathcal{E}^{TF}$, which is defined for all nonnegative ρ in $L^{5/3}(\mathbf{R}^3) \cap L^1(\mathbf{R}^3)$, is called the **Thomas-Fermi energy functional**. It was introduced independently by Thomas [T] and Fermi [FE] shortly after Schrödinger's discovery, but with the constant K replaced by $K^c = 3(3\pi^2)^{2/3}/5$ which, as remarked above, is conjectured to be the sharp constant in Theorem 1. They were, of course, unaware of Theorem 1 which was proved only in 1975, but they proposed that the **Thomas-Fermi ground state energy**

$$(4.18) \quad E^{TF} = E^{TF}(N, z) \equiv \inf \left\{ \mathcal{E}^{TF}(\rho) : \int \rho = N \right\}$$

should be a good approximation to E . We have seen that, apart from minor errors, $E^{TF} \leq E$ when K is used instead of K^c .

The minimization problem (4.18) is an interesting problem in itself and has been analyzed in great detail [LS, L3]. Note that it is defined for all $N \geq 0$, not just integral N . A simple scaling $\rho(x) \rightarrow z^2 \tilde{\rho}(z^{1/3}x)$ replaces z by 1, N by N/z and $E^{TF}(N, z) = z^{7/3} E^{TF}(N/z, 1)$. It turns out that the absolute minimum of $\mathcal{E}^{TF}(\rho)$ without regard to the value of N always occurs when $N = z$, i.e., for the neutral atom. There is a unique minimizing ρ^{TF} for (4.18) if $N \leq z$ and there is no minimizer if $N > z$. In either case

$$(4.19) \quad E^{TF}(N, z) \geq E^{TF}(z, z) = -(\text{const.}) z^{7/3}$$

with $(\text{const.}) = -E^{TF}(1, 1) > 0$.

The remarks in the preceding paragraph apply only to the TF energy and not to the Schrödinger energy but, since $E > E^{TF}$

(with K) to within small errors, we have that the energy of a large atom is bounded below by $-(\text{const.})z^{7/3}$. That this is indeed the correct power law was shown by Lieb and Simon [LS]. See also [L3], where a simpler proof is given.

Theorem 2 (TF Theory is asymptotically exact). *Fix any $0 < \lambda \leq 1$ and let $z_N = N/\lambda$ for $N = 1, 2, 3, \dots$. Let $E(N, z_N)$ be the Schrödinger ground state energy (2.13) for an atom with nuclear charge z_N and N electrons and let $\rho_N(x)$ in (4.5) be the density for any ground state ψ_N . Let $E^{TF}(N, z_N)$ and ρ_N^{TF} be the corresponding Thomas-Fermi energy and density with $K = K^c$ (not the smaller K of Theorem 1). Then*

$$(4.20) \quad \begin{aligned} \lim_{N \rightarrow \infty} E(N, z_N)/E^{TF}(N, z_N) \\ = \lim_{N \rightarrow \infty} E(N, z_N)/[z^{7/3}E^{TF}(\lambda, 1)] = 1. \end{aligned}$$

In the sense of weak $L^1(\mathbf{R}^3)$ convergence on compact sets,

$$(4.21) \quad \lim_{N \rightarrow \infty} z^{-2} \rho_N(z^{-1/3}x) = \lim_{N \rightarrow \infty} z^{-2} \rho_N^{TF}(z^{-1/3}x) = \tilde{\rho}(x)$$

where $\tilde{\rho}$ is the TF minimizer for $z = 1$ and $N = \lambda$.

If $\lambda > 1$ is fixed and $z_N = N/\lambda$ then (4.20) and (4.21) hold with the following replacements:

$$\begin{aligned} E^{TF}(N, z_N) &\rightarrow E^{TF}(z_N, z_N), \\ E^{TF}(\lambda, 1) &\rightarrow E^{TF}(1, 1), \quad \rho_N^{TF} \rightarrow \rho_{z_N}^{TF} \end{aligned}$$

and $\tilde{\rho}$ is the minimizer for $z = 1$ and $\lambda = 1$. In other words, the Schrödinger quantities for an atom with net negative charge converge to the corresponding TF quantities for a neutral atom.

From these considerations we can conclude

(1) *The energy of a large, neutral atom grows with z like $z^{7/3}$. This is significantly different from the earlier result (4.3) without the Pauli exclusion principle because it means that the average energy of each electron is $-z^{4/3}$ instead of $-z^2$.*

(2) *For large z almost all of the electrons are located at a distance $z^{-1/3}$ from the nucleus. The average electron density in this ball of radius $z^{-1/3}$ is z^2 . Without the Pauli exclusion principle we found the distance to be z^{-1} and the electron density to be z^4 .*

Conclusion (2) is that even with the Pauli exclusion principle a large z atom is smaller than a small z atom. This seems to

contradict the experimental fact that “The large range of atomic masses is not accompanied by a correspondingly large or systematic variation in size. Atomic radii all lie between 0.5×10^{-8} cm and 2.5×10^{-8} cm, with no marked increase from the lightest to the heaviest [FR].” While it is not possible to give an unambiguous definition of the radius of something as fuzzy as an atom, it appears [AI] that the largest atom is cesium ($z = 55$) and not uranium ($z = 92$).

The paradox can be resolved by a closer examination of the meaning of “atomic size”. If we define the radius to be the radius of the ball in which most of the electrons reside then $z^{-1/3}$ is, indeed, the correct answer. But that is not the radius that the chemist sees or the distance between atoms in a solid. This “chemical radius” *can be* defined, for example, as the radius R such that

$$(4.22) \quad \int_{|x| > R} \rho_{\psi}(x) dx = \frac{1}{4}$$

for a ground state ψ . In Thomas-Fermi theory this radius is independent of z for large z (because for large $|x|$, $\rho^{TF}(x) = C|x|^{-6}$ when $z = N$, with C independent of z). Theorem 2 (4.21) tells us what ρ_N looks like on distance scales of order $z^{-1/3}$ where ρ_N is of the order z^2 , but it says nothing about distance scales of order one where ρ_N is *presumably* of order one. Nothing is known rigorously about this latter distance scale (and even numerical calculations are uncertain here), but it is surely the case that the true density ρ_N in an atom with $z = N$ and with z large looks like that shown *very* schematically in Figure 1. Also shown in Figure 1 is the innermost region of radius z^{-1} . Thomas-Fermi theory states that $\rho^{TF}(x) = (\text{const.}) z^3$ when $|x| = z^{-1}$ and $\rho^{TF}(0) = \infty$. It is unproved, but undoubtedly true, that when z is large $\rho(x) = Cz^3[1 - O(z|x|)]$ for $0 \leq |x| \leq z^{-1}$, and there is even a conjecture about the precise value of C . The investigation of the inner and outer parts of large atoms is now a subject of active research.

Using poetic license again, we can say:

A large atom is like a galaxy. It has a small, high density, energetic core in which most of the electrons are to be found and whose size decreases with z like $z^{-1/3}$. The electron-electron electrostatic repulsion always manages to push a few electrons out to a distance of order one which is roughly independent of z ; this is the radius one observes if one tries to “touch” an atom.

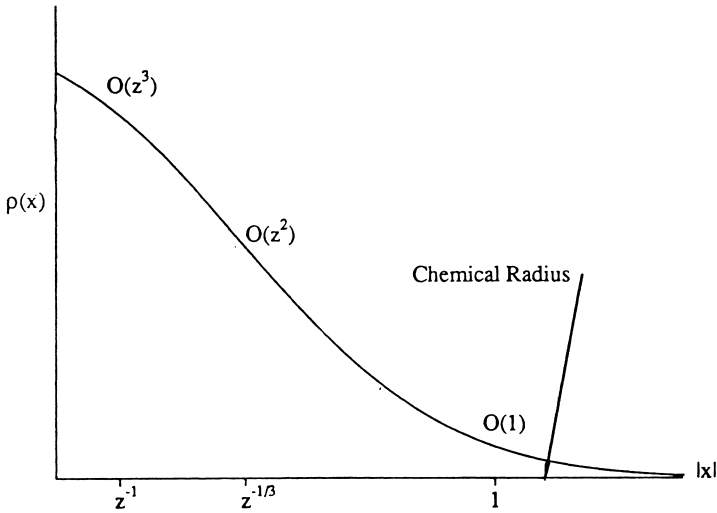


FIGURE 1.

The electron density $\rho(x)$ for a large z atom is crudely plotted as a function of $|x|$, the distance from the nucleus. The graph is intended to show three regimes. (i) At distances of order $z^{-1/3}$, $\rho(x)$ is of order z^2 and most of the electrons are to be found here. (ii) At the very short distance z^{-1} , $\rho(x)$ is of order z^3 . (iii) At distances of order 1, $\rho(x)$ is of order 1 and chemistry takes place. These last two assertions have not yet been established rigorously.

PART V. BULK MATTER: MANY ELECTRONS AND MANY NUCLEI

The analysis given in Part IV can be applied to the general case of N electrons, k nuclei with charges $\underline{Z}$ and locations $\underline{R}$, and with the full potential energy $W(\underline{X})$ in (1.4). The goal is to prove stability of the second kind: $E(N, k, \underline{Z}) \geq -A(z)(N + k)$ for a suitable z -dependent constant $A(z)$, cf. (2.14). This is the theorem first proved by Dyson and Lenard [DL, LE], but with $A(z)$ unrealistically large—of the order $10^{14}z^2$.

Before giving the proof let us discuss the crucial importance of this linear law. For simplicity, assume that all the z_j 's have the same value z and that $N = kz$, i.e., that the system is neutral. First of all, by a suitable choice of a comparison function ψ in

(2.13) on the one hand, and using the linear lower bound plus Newton's theorem on the other hand, Lebowitz and Lieb proved the following [LL].

Theorem 3 (The existence of the thermodynamic limit).

There is a number $A'(z)$ such that

$$\lim_{N \rightarrow \infty} N^{-1} E(N, k = N/z, Z) = -A'(z).$$

The linear law is not just a lower bound it is, indeed, the correct *asymptotic law*. This fact is crucial for the following argument. (There are historical and physical reasons that the words thermodynamic limit instead of asymptotic law are used to describe Theorem 3.)

Suppose now that we have two large collections of N electrons and $k = N/z$ nuclei which are far apart from each other. This means that the nuclear coordinates $\underline{R} = (R_1, \dots, R_{2N})$ satisfy $|R_i - R_{N+j}| > d$ for all $i, j \in \{1, \dots, N\}$ and where d is some very large number. Think of two well separated stones. The ground state energy would then be nearly $2E(N)$ where $E(N) \equiv E(N, k, Z)$. If the two stones are allowed to come together, the ground state energy becomes $E(2N)$ and the energy released is $\delta E = 2E(N) - E(2N)$. With the linear law in the form of Theorem 3, $\delta E = o(N)$; in fact it turns out that $\delta E \sim N^{2/3}$, which is the attractive surface energy that makes the two stones eventually adhere to each other. If, on the other hand, $E(N)$ is proportional to $-N^p$ with $p > 1$ then $\delta E = -(2^p - 2)E(N)$. Thus, the released energy would be of the order of the energy in each stone which, using the actual physical value of $E(N)$, is more than 1000 times larger than the energy released in an explosion of TNT of the same mass. As we shall see in Part VI, hypothetical matter composed solely of bosons would have $E(N) \approx -N^{7/5}$. Using this value of $E(N)$ instead of the energy in a physical stone enhances the energy by a factor $N^{2/5}$; the energy in a smallish stone, $N = 10^{23}$, would then be that of a hydrogen bomb [D]. Such "matter" would be very unpleasant stuff to have lying around the house.

A somewhat less dramatic, but more profound way to state the importance of the thermodynamic limit, Theorem 3, is that it is part of the foundation of thermodynamics [LL]. If this $N \rightarrow \infty$ limit did not exist the argument in the preceding paragraph shows that matter would not behave the way we expect it to behave—even if stability of the second kind is assumed to hold.

Returning to the problem of proving stability of the second kind, we can use Theorem 1 together with the replacement (4.16) (which

yields a controllable error) to deduce that $E(N, k, \underline{Z}, \underline{R}) \geq E^{TF}(N, k, \underline{Z}, \underline{R})$ when E^{TF} is given by (4.18) but now

$$(5.1) \quad \begin{aligned} \mathcal{E}^{TF}(\rho) = & K \int_{\mathbf{R}^3} \rho(x)^{5/3} dx - \int_{\mathbf{R}^3} V(x) \rho(x) dx \\ & + \frac{1}{2} \int_{\mathbf{R}^2} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x - y|^{-1} dx dy + U. \end{aligned}$$

The Thomas-Fermi functional in (5.1) is like that in (4.17) except that $z|x|^{-1}$ is replaced by $V(x)$ and U is added (cf. (1.4)–(1.8)). These depend on $\underline{Z}$ and $\underline{R}$.

A lower bound for E^{TF} can be obtained by omitting the condition that $\int \rho = N$. The absolute minimum occurs [LS, L3] for the neutral case

$$(5.2) \quad N = \sum_{j=1}^k z_j,$$

as it did for the atom in (4.17). One can struggle with minimizing (5.1) or, preferably, one can use a theorem of Teller (see [LS, L3]) which states that *atoms do not bind in Thomas-Fermi theory*. For arbitrary N , this means that there are k positive numbers $N_1, \dots, N_k$ with $\sum_{j=1}^k N_j = N$ such that

$$(5.3) \quad E^{TF}(N, k, \underline{Z}, \underline{R}) \geq \sum_{j=1}^k E_{atom}^{TF}(N_j, z_j),$$

where E_{atom}^{TF} is the solution to the atomic TF minimization problem (4.18). In particular, under the assumption of neutrality (5.2), the optimum choice is $N_j = z_j$ and thus

$$(5.4) \quad E^{TF} \left(\sum_{j=1}^k z_j, k, \underline{Z}, \underline{R} \right) \geq \sum_{j=1}^k E_{atom}^{TF}(z_j, z_j).$$

The right side of (5.4) was already computed in (4.19) and therefore, assuming (4.16),

$$(5.5) \quad E(N, k, \underline{Z}) \geq -(\text{const.}) \sum_{j=1}^k z_j^{7/3}.$$

This is our desired goal of stability of the second kind with $A(z) = (\text{const.})z^{7/3}$. However, errors were made when (4.16) was used. If these are included and bounded in a straightforward way, it turns

out that

$$(5.6) \quad E(N, k, \underline{Z}) \geq -(\text{const.}) \left\{ N + \sum_{j=1}^k z_j^{7/3} \right\},$$

which also implies stability of the second kind. The constant in (5.6) turns out to be about 5 instead of 10^{14} as in [DL].

Incidentally, it should not be inferred that instability of the second kind is necessarily associated with singularities and/or slow fall-off of the potential at infinity – as is the case with the electrostatic potential $|x|^{-1}$. In other words, seemingly “nice” potentials can fail to produce stability of the second kind. Indeed, Thirring [TW, p. 258] shows that if the function $|x|^{-1}$ is replaced everywhere in (1.4)–(1.8) by the seemingly harmless function $(1 + 2|x|) \exp[-|x|]$ then stability of the second kind does not hold for the modified system.

Having proved stability of the second kind with electrostatic forces, we are left with another question which was stated by Ehrenfest in a 1931 address to Pauli (quoted in [D]):

“We take a piece of metal. Or a stone. When we think about it, we are astonished that this quantity of matter should occupy so large a volume. Admittedly, the molecules are packed tightly together, and likewise the atoms within each molecule. But why are the atoms themselves so big?

Consider for example the Bohr model of an atom of lead. Why do so few of the 82 electrons run in the orbits close to the nucleus? The attraction of the 82 positive charges in the nucleus is so strong. Many more of the 82 electrons could be concentrated into the inner orbits, before their mutual repulsion would become too large. What prevents the atom from collapsing in this way? Answer: only the Pauli principle, ‘No two electrons in the same state.’ That is why atoms are so unnecessarily big, and why metal and stone are so bulky.

You must admit, Pauli, that if you would only partially repeal your prohibitions, you could relieve many of our practical worries, for example the traffic problem on our streets.”

One simple measure of the “size” of our many-body system in a state ψ is the average radius $R_p(\psi)$ defined by

$$(5.7) \quad \begin{aligned} [R_p(\psi)]^p &= \frac{1}{N} \int_{\mathbf{R}^3} |x|^p \rho_\psi(x) dx \\ &= \frac{1}{N} \int \sum_{i=1}^N |x_i|^p |\psi(\underline{X})|^2 dx_1 \cdots dx_N. \end{aligned}$$

Here, p is any convenient fixed, positive number. According to our experience with ordinary matter, we should expect $R_p(\psi)$ to grow with N like $N^{1/3}$, i.e., the volume grows like N , when ψ is taken to be a ground state for N electrons. Indeed this surmise is correct—in fact it is correct, as will now be proved, even for some ψ 's which are quite far from being ground states [L1] (see also [LT2]).

Assume that the locations of the nuclei and a normalized ψ satisfy only the condition

$$(5.8) \quad E_\psi \leq 0,$$

which certainly includes the absolute ground states because $E(N, k, \underline{Z}) < 0$ always. Then

$$(5.9) \quad E_\psi = \frac{1}{2}T_\psi + E'_\psi$$

where E'_ψ is the energy in (2.12) but with T_ψ replaced by $\frac{1}{2}T_\psi$ (which amounts to doubling the electron mass). It is easy to see, by a simple scaling $\psi(\underline{X}) \rightarrow 2^{3N/2}\psi(2\underline{X})$ and $R_j \rightarrow \frac{1}{2}R_j$ that $E'(N, k, \underline{Z}, 2\underline{R}) = 2E(N, k, \underline{Z}, \underline{R})$ and therefore $E'(N, k, \underline{Z}) = 2E(N, k, \underline{Z})$. By (5.8) and (5.9) and stability of the second kind

$$(5.10) \quad T_\psi \leq 4|E(N, k, \underline{Z})| \leq 4A(z)(N + k).$$

An easy general inequality is the existence of a positive constant C_p such that

$$(5.11) \quad \left\{ \int_{\mathbf{R}^3} \rho(x)^{5/3} dx \right\} \left\{ \int_{\mathbf{R}^3} |x|^p \rho(x) dx \right\}^{2/p} \geq C_p \left\{ \int_{\mathbf{R}^3} \rho(x) dx \right\}^{5/3+2/p}$$

for all nonnegative ρ . Using the basic kinetic energy inequality (4.9) together with (5.10) and the fact that $\int \rho_\psi = N$, we conclude that

$$(5.12) \quad 4A(z)(N + k) \geq T_\psi \geq KC_p N^{5/3} R_p(\psi)^{-2}.$$

We do not want to impose neutrality (5.2) but let us assume $N \leq 2\sum_j z_j$; one can show [L4] that this is the only case that needs to be treated. Then (5.12) implies that for each $p > 0$

$$(5.13) \quad R_p(\psi) \geq C'_p N^{1/3}$$

as required. Here C'_p depends on p and on z , the maximum of the nuclear charges.

In summary, *bulk matter is stable, and has a volume proportional to the number of particles, because of the Pauli exclusion principle for fermions (i.e., the electrons). Effectively the electrons behave like a fluid with energy density $\rho_\psi^{5/3}$, and this limits the compression caused by the attractive electrostatic forces.*

PART VI. BOSONS IN BULK

The previous paragraph states *inter alia* that the Pauli exclusion principle is essential for stability of the second kind. The verification of this assertion and the dependence of the energy on N without the Pauli principle is itself an interesting mathematical problem. As explained in Part IV, omitting the Pauli principle is equivalent to studying the minimization problem (2.13) for *symmetric* functions, i.e., bosons.

THE $N^{5/3}$ LAW FOR BOSONS. First, let us see what happens if we try to use the Thomas-Fermi energy as a lower bound—as was done in Part V. Theorem 1 is not available for bosons and (4.7) must be used in place of (4.9). The difference between K^1 and K is not important, but the factor $N^{-2/3}$ is crucial. For fixed N , the substitution $K \rightarrow N^{-2/3}K^1$ in (5.1) is mathematically equivalent to replacing the electron mass (which is 1 in our units) by $m' = N^{2/3}K/K^1$. By the scaling $\underline{R} \rightarrow \underline{R}/m'$ and $\psi(\underline{X}) \rightarrow (m')^{3N/2}\psi(m'\underline{X})$ for the Schrödinger theory or $\rho(x) \rightarrow m'\rho(m'x)$ in TF theory one easily finds, in both theories, that $E(N, k, \underline{Z}) \rightarrow m'E(N, k\underline{Z})$. Thus, using (5.6),

$$(6.1) \quad E(N, k, \underline{Z}) \geq -(\text{const.})N^{2/3} \left(N + \sum_{j=1}^k z_j^{7/3} \right)$$

for bosons.

This certainly violates stability of the second kind. If we take $z_j = z$ and $Nk = z$ as before then $E(N) \geq -(\text{const.})N^{5/3}$. Is this result correct or is it simply a bad lower bound? The answer is that *the $N^{5/3}$ law for bosons with fixed nuclei* is correct. It can easily be proved by the following simple choice [L5] of a comparison function ψ for (2.13) and nuclear coordinates $R_1, \dots, R_k$.

Let $f: \mathbf{R}^1 \rightarrow \mathbf{R}$ be given by $f(x) = \sqrt{3/2}[1 - |x|]$ if $|x| \leq 1$ and $f(x) = 0$ otherwise. For each $\lambda > 0$ define $\phi_\lambda: \mathbf{R}^3 \rightarrow \mathbf{R}$ by

$$(6.2) \quad \phi_\lambda(x^1, x^2, x^3) = \lambda^{3/2}f(\lambda x^1)f(\lambda x^2)f(\lambda x^3)$$

with $x = (x^1, x^2, x^3)$, so that for all λ

$$(6.3) \quad \|\phi_\lambda\|_2 = 1 \quad \text{and} \quad T_{\phi_\lambda} = 9\lambda^2.$$

The support of ϕ_λ is the cube $\Gamma_\lambda = \lambda^{-1} \cdot [-1, 1]^3$ of volume $8\lambda^{-3}$. Our comparison function will be

$$(6.4) \quad \psi_\lambda(\underline{X}) = \prod_{i=1}^N \phi_\lambda(x_i),$$

so that

$$(6.5) \quad \rho_{\psi_\lambda}(x) = N\phi_\lambda(x)^2 \quad \text{and} \quad T_{\psi_\lambda} = 9N\lambda^2.$$

For this ψ_λ , (4.16) turns out to be nearly exact.

$$(6.6) \quad \int B(\underline{X}) \psi_\lambda(\underline{X})^2 dx_1 \cdots dx_N \\ = \frac{1}{2} \frac{N-1}{N} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho_{\psi_\lambda}(x) \rho_{\psi_\lambda}(y) |x-y|^{-1} dx dy$$

and the last integral is proportional to N^2/λ . Now assume for simplicity that N/z is an integer and take $k = N/z$ nuclei of charge z . It is possible [L5] to place these nuclei in Γ (roughly in a periodic arrangement) so that the *total* potential energy, $W = -A + B + U$, is equal to $-Cz^{2/3}N^{4/3}\lambda$, where C is some positive constant. (To understand this, note that in a volume which is roughly $d^3 = 1/k\lambda^3$ there is one nucleus and z units of negative charge. The total potential energy is then roughly $-kz^2/d$.) With these favorable locations of the nuclei,

$$(6.7) \quad E_\psi = 9N\lambda^2 - Cz^{2/3}N^{4/3}\lambda.$$

When this is minimized with respect to λ , one finds $\lambda = \frac{C}{18}z^{2/3}N^{1/3}$ and therefore the upper bound

$$(6.8) \quad E(N, k, \underline{Z}) \leq -\frac{C^2}{36}z^{4/3}N^{5/3}.$$

Thus bosonic matter is not only unstable of the second kind but also its radius, λ^{-1} , decreases with N like $N^{-1/3}$. We see here an example of the fact that lack of stability is intimately connected with collapse in the geometric sense.

THE $N^{7/5}$ LAW FOR BOSONS. The $N^{5/3}$ law, which was derived above, is not the end of the story. Observe that it was crucially important that the nuclei were fixed points that could be

located at will in space. Suppose we dispense with that approximation and treat the nuclei also as dynamical quantum mechanical particles. Of course they will also be assumed to be bosons, for if they are fermions then we would merely have a charge reversal of the situation described in Part V, i.e., stability of the second kind would hold.

For simplicity we take $z = 1$ and assume that these positive bosons also have unit mass. The Hamiltonian is then, with x_i denoting the coordinates of the negative bosons and R_i denoting the coordinate of the positive bosons,

$$(6.9) \quad H_{N,k} = - \sum_{i=1}^N \Delta_{x_i} - \sum_{j=1}^k \Delta_{R_j} + W(\underline{X}, \underline{R}).$$

The potential function $W(\underline{X}, \underline{R})$ is that given in (1.4)–(1.8) with $e = 1$; there it was denoted simply by $W(\underline{X})$. The operator $H_{N,k}$ acts on functions $\psi(\underline{X}, \underline{R}) = \psi(x_1, \dots, x_N, R_1, \dots, R_k)$ in $\otimes^N \otimes^k L^2(\mathbf{R}^3)$. The ground state energy is given by (2.13), as usual, with $E_\psi \equiv (\psi, H_{N,k} \psi)$. As explained in Part IV, the optimum choice for ψ , neglecting symmetry considerations, is always a function that is symmetric in the x_i 's and symmetric in the R_i 's. But this is precisely the statement that both kinds of particles are bosons.

It is *much more difficult* to estimate the ground state energy for $H_{N,k}$ in (6.9) than for our previous problem which uses the H in (2.17). The first person to do this rigorously was Dyson [D] who proved the following in 1967.

Theorem 4 (Upper bound for the energy of bosons).

Without loss of generality assume that $N \leq k$. Then the ground state energy, $E(N, k)$, for bosons with the Hamiltonian $H_{N,k}$ satisfies

$$(6.10) \quad E(N, k) \leq -1.32 \times 10^{-6} N^{7/5}$$

for large N . (It is asserted [D] that this constant can be improved substantially.)

An extended comment is in order here. While the $N^{7/5}$ law (6.10) may seem superficially to be only a small improvement over the $N^{5/3}$ law, the conceptual difference is enormous. The proof in Part IV of stability of the second kind for fermions (electrons) and fixed positive particles (nuclei) could be understood solely in terms of what is called semi-classical considerations, i.e., Thomas-Fermi theory. In (5.1) we see that the last three terms on the right

side are simply the classical electrostatic energy of a charged fluid of density $\rho(x)$ interacting with the fixed nuclei. Once we accept the input from quantum mechanics that the kinetic energy is like that of a fluid with energy density $\rho(x)^{5/3}$, the whole energy has a simple classical interpretation—in other words the one-particle density $\rho(x)$ determines the energy with reasonable accuracy.

On the contrary, Theorem 4 cannot be understood this way. In order to achieve an energy as low as $-N^{7/5}$, intricate correlations between the positive and negative particles have to be built into a comparison $\psi(\underline{X}, \underline{R})$, and then the potential energy cannot be expressed solely in terms of ρ_ψ . A delicate balance between potential and kinetic energies is needed and, in the end, it is impossible to think of (6.10) in a simple way—although Dyson [D] does try to give a heuristic explanation of (6.10). Dyson's comparison ψ that leads to (6.10) is very complicated and five pages are needed just to compute E_ψ . This ψ was suggested by work of Bogolubov in 1947 on the superfluidity of liquid helium and it is similar in many ways to the ψ used by Bardeen, Cooper and Schrieffer in their 1957 Nobel prize winning work on superconductivity. Therefore, if the $N^{7/5}$ law is really correct, and not just an upper bound, this kind of highly correlated ψ will be validated in some weak sense as a good approximation to the true ground state.

Two decades later the requisite lower bound was proved by Conlon, Lieb and H-T. Yau [CLY]. The proof is too involved to explain here, even heuristically, but the main result is the following.

Theorem 5 (Lower bound for the energy of bosons).

Without loss of generality assume that $N \leq k$. Then the ground state energy $E(N, k)$ for bosons with the Hamiltonian $H_{N, k}$ satisfies

$$(6.11) \quad E(N, k) \geq -AN^{7/5}$$

for some universal constant A . If $N = k$ (neutrality) and N is large then A can be taken to be 0.79 in (6.11).

PART VII. RELATIVISTIC MATTER

According to Einstein's 1905 special theory of relativity, the **relativistic kinetic energy** as a function of the momentum $p \in \mathbf{R}^3$ should be

$$(7.1) \quad T(p) = \sqrt{p^2 c^2 + m^2 c^4} - mc^2$$

instead of $p^2/2m$ as in (1.11). Here c is the speed of light. The quantity mc^2 is called the **rest energy** and it has been subtracted off in (7.1). If $p^2 \ll mc^2$ then $T(p) \approx p^2/2m$ as before. Using the first part of (1.13) the relation between p and the velocity v is

$$(7.2) \quad v = pc^2(p^2c^2 + m^2c^4)^{-1/2}.$$

When $p^2 \ll mc^2$ this gives $v \approx p/m$ as before, but if $p^2 \rightarrow \infty$ then $v^2 \rightarrow c^2$. Thus no particle can move faster than the speed of light, and a measure of the importance of relativistic effects is the ratio $|v|/c$.

To gain an understanding of the possible effect of relativistic mechanics on atoms it is convenient to introduce the number

$$(7.3) \quad \alpha = e^2/\hbar c.$$

This is a *dimensionless* constant whose numerical value is $1/137$ and it is called the **fine structure constant**. (The reason for this appellation is that α also governs certain small effects in atoms that are connected with the electron spin and magnetism, and that have not been discussed here; these effects cause a small splitting of each spectral line into several nearby lines—that experimentalists call the fine structure of the line.)

What is the value of $|v|/c$ for an electron in a hydrogenic atom? We can take v^2 in a state ψ to be $(2/m)$ times the kinetic energy T_ψ (which is $p^2/2m$ classically). Using the ground state ψ given by (3.6), and restoring all units (recall that we took $\hbar^2/2m = 1$ and $e = 1$), one easily finds from the definition (2.6) that

$$(7.4) \quad |v|/c = \alpha z.$$

For hydrogen ($z = 1$), $|v|/c = 1/137$ so relativistic effects are unimportant. But when z is large (7.4) correctly gives $|v|/c$ for the “innermost” electrons, and we see that it is not small.

The problem of combining relativistic mechanics with quantum mechanics is an old and extremely difficult one. It is not even easy at the classical mechanics level. At that level one *could* do the following. Return to the classical Hamiltonian function in (1.11) and simply replace $p_i^2/2m$ by $T(p_i)$ and then use the equations of motion (1.13). While such a dynamical theory makes sense mathematically, it is *not* a relativistic theory. The reason is that it is not invariant under Lorentz transformations of space-time unless $W \equiv 0$. To remedy this defect it is necessary to give up the idea of particles interacting instantaneously by a force (1.9) which

depends only on their locations in $\mathbf{R}^3$. Instead, it is necessary to invoke the intervention of the full electromagnetic field, to introduce a piece of the Hamiltonian for this field (whose equations of motion in the spirit of (1.13) turn out to be Maxwell's equations), and to introduce another piece of the Hamiltonian which gives the interaction of the particles with the field. All this is very complicated and the final equations of motion do not even make strict mathematical sense for point particles.

The situation is even worse when quantum mechanics is introduced because then the electromagnetic field also has to be "quantized." This is the enormous subject of *quantum electrodynamics* (Q.E.D) which, in turn, is part of an even larger subject—*quantum gauge field theory*. (Then there is superstring theory which is still more complicated.)

When $\alpha = 0$ (equivalently $e = 0$) Q.E.D. is trivial because then the electromagnetic field is decoupled from the particles and there are no interactions. This suggests that one can solve problems in Q.E.D. by making a power series expansion in α since α is small. Indeed, such a "perturbation theory" has been investigated in great detail and many of its predictions are confirmed incredibly well by experiments. At present, however, nobody knows in what sense this power series converges, if at all, or how to find any except the first few terms in the series, or what a "nonperturbative" theory would predict. In particular, what happens when there is a very large number of electrons and nuclei, in which case a perturbative treatment is inadequate? Does Q.E.D. predict the stability of matter?

It should be stated that the particle aspect of Q.E.D. is not built on the Schrödinger $p^2 = -\hbar^2 \Delta$, which is a second order elliptic operator, but instead on the *Dirac operator* which is a quartet of first order operators which acts on four-component spinors. The Dirac operator is relativistic (i.e., it is invariant under Lorentz transformations), but if we simply add the potential energy term $W(\underline{X})$ to it the theory will still not be relativistic for the same reason as before (i.e., instantaneous interactions are not Lorentz invariant). Nevertheless, this kind of "mixed" theory is often used instead of the Schrödinger Hamiltonian because it presumably takes account of most of the relativistic corrections. However, from our point of view the Dirac Hamiltonian has a much more serious defect: the ground state energy is *not* given by a minimization problem as in (2.13). In fact the spectrum of the Dirac Hamiltonian is not bounded below and Dirac had to introduce an extra axiom in order to decide which of the many eigenvalues should be considered as

the ground state energy. (This axiom is known as “filling the negative energy sea;” unfortunately, it is ambiguous in the many-body case. The axiom also led to the prediction of certain kinds of elementary particles called *antimatter*; the first one to be discovered was the positron (antielectron) by Anderson in 1932.)

In order to investigate the stability of matter with relativistic kinetic energy, $T(p)$, in a mathematically rigorous way, we are led to study the following theory which uses a “relativistic” modification of the Schrödinger energy. This theory is a caricature of the proper theory, but it will have the advantage of being a well-posed minimization problem as before, and one which is tractable—unlike Q.E.D. At present it is the best that can be done to analyze the stability of matter question rigorously.

With ψ as before, replace the kinetic energy of (2.11) by

$$(7.5) \quad T_\psi = \sum_{i=1}^N (\psi, T_{op}(p_i) \psi)$$

where $T_{op}(p)$ is given by (7.1), but with $p_i = -i\hbar\nabla_{x_i}$ as before. In other words, T_{op} is the operator

$$(7.6) \quad T_{op} = \sqrt{-\hbar^2 c^2 \Delta + m^2 c^4} - mc^2$$

which is perfectly respectable although, unlike the operator $-\hbar^2 \Delta / 2m$, it is not local, i.e., $(T_{op}\psi)(x)$ is *not* determined by ψ in an infinitesimally small neighborhood of x . In terms of the Fourier transform $\hat{\psi}$ given by (2.4) for one particle,

$$(7.7) \quad T_\psi = \int_{\mathbf{R}^3} \left\{ \sqrt{p^2 c^2 + m^2 c^4} - mc^2 \right\} |\hat{\psi}(p)|^2 dp$$

replaces (2.6). The definitions (2.12)–(2.14) for the energy are unaltered.

For the purpose of investigating stability, a simplification can be made, and will be made in this part but not in the next. Since

$$(7.8) \quad c|p| - mc^2 \leq T(p) \leq c|p|$$

the difference of the operators T_{op} and $c\hbar\sqrt{-\Delta}$ is a bounded operator. Therefore, for *both* stability of the first and second kind it suffices to replace T_{op} by the operator $c\hbar\sqrt{-\Delta}$, i.e., we can set $m = 0$. Then T_ψ becomes, for one particle,

$$(7.8) \quad T_\psi = c \int_{\mathbf{R}^3} |p| |\hat{\psi}(p)|^2 dp.$$

The advantage of this replacement is a scaling invariance that will become apparent in (7.14) and (7.15). It also brings into focus the essential feature of any theory of relativistic quantum mechanics (Q.E.D. or the Dirac equation or the Klein-Gordon equation or whatever) which can be stated as follows. The central fact about quantum mechanics is that p^2 becomes the operator $-\hbar^2\Delta$ and this can be thought of heuristically (as in (3.5)) as the reciprocal of a length squared—the length being essentially the width of a function ψ . In nonrelativistic quantum mechanics this operator is also the kinetic energy and it handily controls the potential energy which is proportional to the reciprocal of the same length. In relativistic quantum mechanics, on the other hand, the kinetic energy is essentially $c\sqrt{-\hbar^2\Delta}$ and this is only the reciprocal of the length—not the length squared. Thus, both the potential and kinetic energies are on the same footing in a relativistic quantum theory, and we therefore have what would usually be termed “the critical case”.

A word about constants is required here. In the nonrelativistic case we used scaling, as stated at the beginning of Part III, to eliminate all constants except for the nuclear charge numbers z_j . This cannot be done now because T_ψ and W_ψ scale in the same way. We can, however, use units in which $\hbar c = 1$. Then the energy becomes

$$(7.9) \quad E_\psi = T_\psi + \alpha W_\psi$$

where W_ψ is given by (2.10) and (1.4) with e^2 set equal to one there. T_ψ is given by

$$(7.10) \quad T_\psi = \sum_{i=1}^N \left(\psi, \sqrt{-\Delta_{x_i}} \psi \right) = \int_{\mathbf{R}^{3N}} \sum_{i=1}^N |p_i| |\hat{\psi}(p_1, \dots, p_N)|^2 dp_1 \cdots dp_N$$

with $\hat{\psi}$ being the $\mathbf{R}^{3N}$ Fourier transform. The Hamiltonian that replaces (2.17) is

$$(7.11) \quad H = \sum_{i=1}^N \sqrt{-\Delta_{x_i}} + \alpha W(\underline{X}).$$

Thus there are now two constants in the problem: the fine structure constant α and z , the maximum nuclear charge.

Let us begin by analyzing the **hydrogenic atom**. In this case only the combination $z\alpha$ enters because $U = 0$. Heuristic considerations, like those in Part III, would lead us to investigate (cf. (3.5))

$$(7.12) \quad \min_L \left\{ \frac{1}{L} - \frac{z\alpha}{L} \right\}$$

as an approximation to E . This quantity is zero if $z\alpha \leq 1$ and $-\infty$ if $z\alpha > 1$. Indeed this conclusion is qualitatively correct because of an inequality of Kato [K] and Herbst [H] (see also [LY1])

$$(7.13) \quad \int_{\mathbf{R}^3} |p| |\hat{\psi}(p)|^2 dp \geq \frac{2}{\pi} \int_{\mathbf{R}^3} |x|^{-1} |\psi(x)|^2 dx$$

in which $2/\pi$ is the sharp constant. Consequently, the hydrogenic atom has the following ground state energy

$$(7.14) \quad \begin{aligned} E &= 0, & \text{if } z\alpha \leq \frac{2}{\pi} \\ &= -\infty & \text{if } z\alpha > \frac{2}{\pi} \end{aligned}$$

and *stability of the first kind holds if and only if* $z\alpha \leq 2/\pi$, i.e., $z \leq 87$ with $\alpha = 1/137$.

Relativistic quantum mechanics stabilizes an atom only if $z\alpha$ is small enough.

It is to be noted that a similar catastrophe occurs with the Dirac Hamiltonian for an atom [KS]. Again, only $z\alpha$ enters and there is a critical value $z\alpha = 1$, instead of $2/\pi$. For $z\alpha \leq \sqrt{3}/2$ this operator is essentially selfadjoint on C_c^∞ , the infinitely differentiable functions of compact support. When $z\alpha > \sqrt{3}/2$ the deficiency indices are $(+1, +1)$ and so there is a selfadjoint extension, but only for $\sqrt{3}/2 < z\alpha < 1$ is there a *distinguished*, physical selfadjoint extension. It is distinguished either by analytic continuation from the $z\alpha \leq \sqrt{3}/2$ case or by finiteness of the potential energy; in other words, when $z\alpha \geq 1$ the kinetic and potential energies cannot be defined separately. When $z\alpha = 1$ the hydrogenic ground state energy “falls into the negative energy sea” and the Dirac operator then ceases to make good physical sense. The underlying reason for the catastrophe is the same in both cases: the kinetic energy is $1/L$ instead of $1/L^2$.

The next step is to investigate **large atoms**, as in Part VI. Clearly, *stability of the first kind holds here if and only if* $z\alpha \leq 2/\pi$ because, as a simple comparison function $\psi(\underline{X})$ demonstrates, the positive electron-electron repulsion $B(\underline{X})$ of (1.7) cannot overcome the $-\infty$ energy to be gained by letting one electron “fall into the nucleus” when $z\alpha > 2/\pi$.

In fact conclusion (7.14) holds in the large atom case as well. This is where the nice scaling property of $\sqrt{-\Delta}$ and $|x|^{-1}$ come in. To jump ahead for the moment, we can always say, when $m = 0$, that in the *general case* of many electrons and many nuclei there are precisely two alternatives:

$$(7.15) \quad \text{either } E(N, k, \underline{Z}) = 0 \quad \text{or} \quad E(N, k, \underline{Z}) = -\infty,$$

which tells us that *stability of the first kind is equivalent to stability of the second kind for relativistic matter*. The proof of (7.15) is a simple one: If $E_\psi < 0$ for some ψ then, by setting $\psi_\lambda(\underline{X}) = \lambda^{-3N/2} \psi(\lambda \underline{X})$, we have that $E_{\psi_\lambda} = \lambda E_\psi$, and this can be driven to $-\infty$ by letting $\lambda \rightarrow \infty$. If, on the other hand, $E_\psi > 0$ then we can drive E_{ψ_λ} to 0 by letting $\lambda \rightarrow 0$.

Apart from the stability question we can also ask about the *structure* of large atoms—as in Part IV. Is there an appropriate Thomas-Fermi theory in this case? The answer is *No* and it is instructive to understand why.

We can start by asking for an analogue of Theorem 1, which is the basic kinetic energy estimate for fermions. There is one—as discovered by Daubechies [DA]: *Theorem 1 holds for the $\sqrt{-\Delta}$ kinetic energy if $5/3$ is replaced everywhere by $4/3$, if $2^{2/3}$ is replaced by $2^{1/3}$ in (4.8) and if K is changed. ($K = 1.26$ will do.)*

So far, so good. Next we form the Thomas-Fermi functional as in (4.17):

$$(7.16) \quad \begin{aligned} \mathcal{E}^{TF}(\rho) = & K \int_{\mathbf{R}^3} \rho(x)^{4/3} dx - z\alpha \int_{\mathbf{R}^3} |x|^{-1} \rho(x) dx \\ & + \frac{1}{2} \alpha \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x - y|^{-1} dx dy. \end{aligned}$$

This, unfortunately, is unbounded below for any choice of $N = \int \rho$. (The reader can easily verify that $\int \rho^{4/3}$ cannot control the $|x|^{-1}$ singularity, and the last term in (7.16) does not save the situation.) The conclusion is that Thomas-Fermi theory is useless for large atoms (and hence also for the case of many electrons and many nuclei), but in Part VIII we shall see that it *is* useful for describing the relativistic matter interacting by gravitational forces in a star.

Finally, we turn to **bulk matter** as in Part V. Now, however, even stability of the first kind is problematic for one electron and many nuclei; it already is problematic even for one electron ($N = 1$) and one nucleus ($k = 1$). The following additional concern now

presents itself. Suppose there are two nuclei with $z\alpha = 3/2\pi$, located at R_1 and R_2 in $\mathbf{R}^3$, and suppose there is but one electron. If the two points are coincident, say $R_1 = R_2 = 0$, we have a “bomb.” While each nucleus is subcritical the combined nucleus is supercritical, $\tilde{z}\alpha = 3/\pi$, and the electron can fall into it, releasing an infinite energy (7.15). What prevents this from happening? Answer: the nucleus-nucleus repulsive energy

$$(7.17) \quad U = \alpha z^2 |R_1 - R_2|^{-1} = \frac{(z\alpha)^2}{\alpha} |R_1 - R_2|^{-1},$$

which goes to $+\infty$ as $R_1 \rightarrow R_2$ and which is one part of $E(1, 2, \underline{Z}, \underline{R})$. There are now genuinely *two* parameters in the problem, which can conveniently be taken to be $z\alpha$ and α . By length scaling, all the energies are proportional to $|R_1 - R_2|^{-1}$, so $E(1, 2, \underline{Z}, \underline{R}) = C|R_1 - R_2|^{-1}$. If $z\alpha$ is fixed at $3/2\pi$, is the constant $1/\alpha$ in (7.16) large enough so that the repulsion U wins, i.e., so that $C \geq 0$?

This exercise with $N = 1$ and $k = 2$ tells us that stability will require not only an upper bound on $z_j\alpha$ of $2/\pi$ for each nucleus but *also* an upper bound, α_c , for α . This is clarified in Theorem 6 below. Furthermore, we might fear that this α_c goes to zero as $z\alpha \rightarrow 2/\pi$. Here, $z = \max_j \{z_j\}$ as before. It might also happen that α_c depends on N and/or k . Fortunately, neither of these two fears materializes.

The first result on this problem was by Daubechies and Lieb [DAL] who proved stability for one electron and arbitrarily many nuclei, namely,

$$(7.18) \quad E(1, k, \underline{Z}) = 0$$

provided $z_j\alpha \leq 2/\pi$ for each j and provided $\alpha \leq 1/3\pi$.

The big breakthrough was by Conlon [CO] who proved for all N and k that

$$(7.19) \quad E(N, k, \underline{Z}) = 0$$

provided $z_j = 1$ for all j and provided $\alpha \leq 10^{-200}$ [sic] and provided that spin is omitted, i.e., $\psi \in \bigwedge^N L^2(\mathbf{R}^3)$ instead of $\psi \in \bigwedge^N L^2(\mathbf{R}^3; \mathbf{C}^2)$. This was vastly improved by Fefferman and de la Llave [FD] to $\alpha \leq 1/2.06\pi$ with the other conditions remaining the same.

The results of Conlon, Fefferman and de la Llave fall short of the critical case $z\alpha = 2/\pi$. They also do not include all ψ 's because the inclusion of spin is not as easy as it was in the nonrelativistic

case. The matter was finally settled by Lieb and H-T. Yau [LY1] who treated a slightly more general problem. The method of proof in [LY1] is very different from that in [C] and [FD].

Theorem 6 (Stability of relativistic matter up to $z\alpha = 2/\pi$). *Let N and k be arbitrary and let $z_j\alpha \leq 2/\pi$ for each $j = 1, \dots, k$. Let $E_q(N, k, \underline{Z})$ be the infimum of the energy (7.9) as in (2.11) and (2.14), but with $\psi \in \bigwedge_1^N L^2(\mathbf{R}^3; \mathbf{C}^q)$. ($q = 1$ is the simple antisymmetric case, $q = 2$ is the physical case and $q > 2$ is for fun.) Then $E_q(N, k, \underline{Z}) = 0$ if $q\alpha \leq 1/47$. In particular, stability holds in the physical case: $\alpha = 1/137$.*

Is this limitation on α (especially the $1/q$ dependence) merely an artifact of the proof in [LY1]? What about the stability of relativistic bosonic matter? These questions are answered in the following two theorems [LY1]. Roughly, the proof of Theorem 7 uses the same elementary ideas as in the proof of (6.7) and (6.8).

Theorem 7 (Instability of bulk matter for large α). *With the definition of E_q as in Theorem 6, assume there are N electrons and k nuclei, each with the same charge $z > 0$.*

(1) **q and z independent bound.** *If $\alpha > 128/15\pi$, if $N \geq 1$ and if $k \geq 1 + 1/z$ then there is collapse for all q , i.e., $E_q(N, k, \underline{Z}) = -\infty$.*

(2) **q and z dependent bound.** *If $\alpha > 36q^{-1/3}z^{2/3}$, if $N \geq q$ and if $k \geq q/z$ then $E_q(N, k, \underline{Z}) = -\infty$.*

Thus, if α is large one electron can cause collapse no matter how small z is, provided enough nuclei are used in order to make the system approximately neutral. When q is large the critical α decreases with q at least as fast as $q^{-1/3}$. This contrasts with the q^{-1} dependence of Theorem 6; the exact dependence of the critical α on q is not known.

One reason for mentioning the q dependence is that the choice $q = N$ is equivalent to omitting the Pauli exclusion principle altogether. The interested reader can easily deduce this from the discussion of symmetry in Part IV. This means that the case $q = N$ includes the boson case, and as a corollary of Theorem 7 we have

Theorem 8 (Instability of relativistic bosons with fixed nuclei). *Let $\alpha > 0$ and assume that all nuclei have a common charge $z > 0$. If the dynamic, negative particles are bosons instead of fermions, and if the nuclei are fixed as before, then collapse always occurs for sufficiently large N and k , i.e., $E(N, k, \underline{Z}) = -\infty$. The choice $N \geq (36)^3 z^2 \alpha^{-3}$ and $k \geq (36)^3 z \alpha^{-3}$ suffices.*

PART VIII. THE STABILITY AND INSTABILITY OF COLD STARS

Stars shine because they are a vast, continuous nuclear explosion. Nevertheless, the effect of attractive gravitational forces in such a huge body is not negligible—a fact that can easily be appreciated by noting that the sun maintains a spherical shape despite the nuclear violence. The byproducts of the nuclear reactions are light, heat and various kinds of particles, all of which produce an internal pressure that keeps the star expanded before these byproducts can leak out through the surface.

After the nuclear reactions are finished, in what can be called the post-twinkle phase, the star eventually cools, and it might be supposed that it would then collapse because of the gravitational attraction. Sometimes this does happen with great rapidity—for that is what a supernova is; the enormous gravitational energy is then converted into the production of a vast amount of light and numerous particles, mostly neutrinos. But sometimes the cold star, or the remnant of a supernova explosion, which is also a cold star, merely attains a new, much smaller radius and is quite stable—although lightless. (Another conceivable final state of a supernova is a black hole, in which case the collapse might be silent because black holes do not permit light to escape; I thank F. Dyson for this observation.)

The determining factor for total collapse or stability is the stellar mass; the dividing line is several solar masses. What mechanism is available, in the stable case, to provide the internal pressure that prevents the gravitational attraction from crushing the cold star? Answer: *The quantum mechanical kinetic energy of fermions.*

Actually there are two kinds of cold stars: **neutron stars** and **white dwarfs**. The latter kind will be discussed at the end. The former kind, which are remnants of supernovae, are composed mostly of chargeless neutrons. (In Part I it was stated that a free neutron decays into a proton and an electron in about 13 minutes, but in the very dense interior of such a star the decay ceases for several reasons.) With $\underline{X} = (x_1, \dots, x_N)$ denoting the coordinates of N neutrons, the classical potential energy is given by Newton's gravitational formula (cf. (1.2) and (1.9))

$$(8.1) \quad W(\underline{X}) = -\kappa \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}.$$

Here, $\kappa = GM_n^2$, with G being Newton's gravitational constant and M_n being the neutron mass (which is just a little bigger than

the proton mass, M_p). There is no electrostatic potential energy since the neutron has no electric charge.

Neutron stars were mentioned as a possibility by Landau in 1932, after the discovery of the neutron by Chadwick in the same year [B, ST]. In 1934 Baade and Zwicky proposed their connection with supernovae. Their actual existence was "established" by the 1968 discovery of pulsars (which mostly emit accurately timed pulses of radio waves) and Gold's identification of them as rotating neutron stars. The supposed facts about these stars are amazing [B, ST]. Their mass is a few solar masses (about 10^{57} neutrons) but their radius is only about 10 km. Thus the gravitational force at their surface is about 10^{12} times that on Earth. The internal structure of such an object is hard to guess, because at these densities—especially the central density—neutrons will cease to look like individual particles and might become some sort of quark soup. Moreover, this large mass in such a small volume will cause space to be "curved" in accordance with the principles of Einstein's general theory of relativity. These complications will be ignored here and we shall suppose that a neutron star is just a collection of N particles with the gravitational potential energy function of (8.1).

The model to be discussed here is standard and it is built along the lines of Part VII. With an abuse of history we shall call it the Chandrasekhar model, for a reason that will become clear at the end. Since the neutron star is cold we can suppose that it is in its quantum mechanical ground state. The energy of a wave function ψ is given, as usual, by $E_\psi = T_\psi + W_\psi$ with T_ψ defined in (7.5) with $m = M_n$ and W_ψ defined in (2.10) with $W(\underline{X})$ given by (8.1). Adopting henceforth units in which $\hbar = c = 1$, the Hamiltonian is

$$(8.2) \quad H_N = \sum_{i=1}^N \left\{ \sqrt{-\Delta_{x_i} + m^2} - m \right\} - \kappa \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}.$$

with $\kappa = GM_n^2$ and $m = M_n$. The problem is to compute

$$(8.3) \quad E(N) = \inf_{\psi} E_\psi = \inf_{\psi} (\psi, H_N \psi)$$

and to delineate the properties of the density ρ_ψ for a minimizing ψ .

Two things are to be noticed. One is that the neutrons are being treated (special) relativistically with $T(p) = \sqrt{p^2 c^2 + m^2 c^4} - mc^2$. This is obviously important to do since $|v|/c$ will turn out

to be quite large at these densities. As in Part VII, the question of collapse or no collapse can be decided by the simpler choice $T(p) = c|p|$, but if we want to compute the density ρ_ψ in the stable phase it is essential to keep the full expression (8.2).

The second point is that $W(\underline{X})$ has the same form as the electron-electron repulsion term $B(\underline{X})$ in (1.7) except for one thing—the *minus sign*. For the repulsive case we were able to approximate this energy in terms of ρ_ψ as in (4.16), because, as discussed there, the left side of (4.16) minus the right side is bounded *below* [LO] by— $(1.68) \int \rho_\psi^{4/3}$. But because of the minus sign in $W(\underline{X})$ we now require an *upper* bound for the difference in (4.16) in order to achieve a lower bound to E_ψ . Unfortunately the only upper bound is $+\infty$. Therefore, controlling W_ψ by T_ψ is not at all a simple matter in this case.

Nevertheless, there is a simple, well-defined Thomas-Fermi approximation to this problem, as in Part IV, that is obtained using (4.16). Whether it is correct or not remains to be seen. Its construction requires approximating the relativistic kinetic energy (7.5) in terms of ρ_ψ in analogy with Theorem 1 or Theorem 2. Imitating the discussion in Remark 1 after (4.9), we place n particles in a cube of size L and compute the minimum of T_ψ . With $\rho = n/L^3$ and $\eta \equiv (6\pi^2 \rho/q)^{1/3}$ it turns out to be (for large n) equal to the volume L^3 times the quantity

$$(8.4) \quad j(\rho) = \frac{q}{2\pi^2} \int_0^\eta \{(p^2 + m^2)^{1/2} - m\} p^2 dp.$$

(Recall that $q = 1$ for $\psi \in \bigwedge^N L^2(\mathbf{R}^3)$ and $q = 2$ for the physical case $\psi \in \bigwedge_1^N L^2(\mathbf{R}^3; \mathbf{C}^2)$.) This $j(\rho)$ reduces to $(2/q)^{2/3} K^c \rho^{5/3}$ when ρ is small (corresponding to the nonrelativistic kinetic energy) and to $(3(6\pi^2)^{1/3}/4q^{1/3}) \rho^{4/3}$ when ρ is large (corresponding to the relativistic case). It should be remarked that Daubechies [DA], who extended Theorem 1 to $\sqrt{-\Delta}$, as noted before (7.16), also extended it to $\sqrt{-\Delta + m^2} - m$.

We can now form the Thomas-Fermi functional for a neutron star.

$$(8.5) \quad \mathcal{E}^{TF}(\rho) = \int_{\mathbf{R}^3} j(\rho(x)) dx - \frac{\kappa}{2} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x-y|^{-1} dx dy$$

and the Thomas-Fermi energy is

$$(8.6) \quad E^{TF}(N) = \inf \left\{ \mathcal{E}(\rho) : \int \rho = N \right\}.$$

This problem can also be conveniently reformulated by setting

$$(8.7) \quad \tilde{\rho}(x) = \rho(N^{1/3}x) \Rightarrow \int_{\mathbf{R}^3} \tilde{\rho}(x) dx = 1$$

so that, with

$$(8.8) \quad \tilde{\mathcal{E}}^{TF}(\rho) = \int_{\mathbf{R}^3} j(\rho(x)) dx - \frac{1}{2} \kappa N^{2/3} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x-y|^{-1} dx dy,$$

the energy becomes

$$(8.9) \quad \frac{1}{N} E^{TF}(N) = \inf \left\{ \tilde{\mathcal{E}}^{TF}(\rho) : \int \rho = 1 \right\} \equiv e^{TF}(\kappa N^{2/3}).$$

The advantage of (8.7)–(8.9) is that it makes clear that the relevant parameter in the problem is $\tau = \kappa N^{2/3}$. Numerically, $\kappa = Gm_N^2$ is about 10^{-38} in our units and $N \approx 10^{57}$, so that $\tau \approx 1$. This suggests that it is an excellent approximation, indeed, to consider the limit $\kappa \rightarrow 0$, $N \rightarrow \infty$ but with $\kappa N^{2/3} = \tau$ held fixed.

The first attempt [LT3] relate $E(N)$ to $E^{TF}(N)$ succeeded to within a factor of 4. The following theorem of Lieb and H-T. Yau [LY2] finally settled the question.

Theorem 9 (Thomas-Fermi theory is asymptotically exact for cold stars). Fix $\tau = \kappa N^{2/3}$ and let $\kappa \rightarrow 0$ and $N \rightarrow \infty$. Then

$$(8.10) \quad \lim_{\kappa \rightarrow 0, N \rightarrow \infty} E(N)/N = e^{TF}(\tau).$$

(The error, for finite N , can be bounded.) Furthermore, there is a critical number, τ_c , such that the Thomas-Fermi minimization problem (8.8), (8.9) satisfies

(1) If $\tau < \tau_c$ then $e^{TF}(\tau)$ is finite and there is a unique minimizing ρ_τ^{TF} .

(2) If $0 < \tau < \tau_c$ and if ψ_N is a minimizer for the quantum mechanical problem with density ρ_{ψ_N} then, weakly in $L^{4/3}(\mathbf{R}^3) \cap L^1(\mathbf{R}^3)$,

$$(8.11) \quad \lim_{\kappa \rightarrow 0, N \rightarrow \infty} \rho_{\psi_N}(N^{1/3}x) = \rho_\tau^{TF}(x).$$

(3) The function $N \mapsto E^{TF}(N)$, for fixed κ , is concave and decreasing.

(4) If $\tau > \tau_c$ then $e^{TF}(\tau) = -\infty$, while $e^{TF}(\tau_c)$ is finite.

Thus, a neutron star is stable if $\kappa N^{2/3} < \tau_c$ and it collapses if $\kappa N^{2/3} > \tau_c$. It is not hard to see from (8.5) that such a collapse

occurs and, indeed, to infer the value of τ_c . For large ρ (which is relevant here) $j(\rho) = C\rho^{4/3}$ with $C = 3(6\pi^2)^{1/3}/4q^{1/3}$. The sharp inequality

$$(8.12) \quad \begin{aligned} A \left\{ \int_{\mathbf{R}^3} \rho(x) dx \right\}^{2/3} \int_{\mathbf{R}^3} \rho(x)^{4/3} dx \\ \geq \frac{1}{2} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x - y|^{-1} dx dy \end{aligned}$$

is well known, with $A = 1.092$. Therefore

$$(8.13) \quad \tau_c = C/A.$$

The Euler-Lagrange equation for the problem (8.8), (8.9) is

$$(8.14) \quad j'(\rho(x)) = [\eta(x)^2 + m^2]^{1/2} - m = \{\tau|x|^{-1} * \rho - \mu\}_+,$$

where $\eta(x) = (6\pi^2\rho(x)/q)^{1/3}$, $*$ denotes convolution, $\mu > 0$ is a Lagrange multiplier (which has to be adjusted so that $\int \rho = 1$) and $\{A\}_+ = \max(A, 0)$. The optimum solution ρ is spherically symmetric, decreasing and has compact support in a ball B_R of radius R . The Euler-Lagrange equation can be converted into a differential equation by applying the Laplacian, Δ , to (8.14).

$$(8.15) \quad -\Delta\Theta = 4\pi[\Theta^2 + 2\Theta]^{3/2} \quad \text{in } B_R$$

with $\Theta = \kappa|x|^{-1} * \rho$ and $\Theta(x) = 0$ when $|x| = R$.

The existence of a solution to (8.14), (8.15) was proved by Auchmuty and Beals [AB]. The analysis of this equation is, in itself, an interesting mathematical problem. The following has been proved [LY2].

Theorem 10 (Properties of the Thomas-Fermi density). *There is exactly one nonnegative radial solution to (8.14) for each $\mu > 0$ and $0 < \tau < \tau_c$. It has compact support. For $\tau \geq \tau_c$ there is no solution (even though $e^{TF}(\tau_c)$ is finite). If $\tau < \tau_c$ there is exactly one μ so that $\int \rho = 1$. Both ρ and μ are continuous functions of τ and $\rho(0)$ is an increasing function of τ , while the radius R is a decreasing function with $R \rightarrow 0$ as $\tau \rightarrow \tau_c$ and $R \rightarrow \infty$ as $\tau \rightarrow 0$. If $\tau_1 < \tau_2 < \tau_c$ then the solutions ρ_1 and ρ_2 satisfy $\rho_1(x) = \rho_2(x)$ for exactly one value of $|x|$.*

These properties show that the star has a unique shape that evolves continuously with τ , i.e., there are no jumps.

What would happen if the neutrons, which are fermions, are replaced by **bosons**? Elementary particles that are bosons, and that are stable have never been seen, as stated earlier, but they

are posited to exist in certain theories in order to account for the “missing mass” in the universe. They are called axions and are not expected to condense into stars because they cannot get rid of their gravitational energy easily, but it is amusing to consider the possibility anyway. See [RB].

For bosons we might imagine that a simple product function $\psi(\underline{X}) = \Pi \phi(x_i)$ with $\phi \in L^2(\mathbf{R}^3)$, as in (4.2), would adequately describe the ground state. Then $\rho_\psi(x) = N|\phi(x)|^2$ and with $\phi(x) \geq 0$ (which is the optimum choice),

$$(8.16) \quad T_\psi = N(\phi, T_{op}\phi) = \left(\sqrt{\rho_\psi}, T_{op}\sqrt{\rho_\psi} \right)$$

$$(8.17) \quad W_\psi = \frac{\kappa}{2} \frac{N-1}{N} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho_\psi(x) \rho_\psi(y) |x-y|^{-1} dx dy.$$

Adding these (and supposing N to be large) we obtain a **Hartree type functional**

$$(8.18) \quad \mathcal{E}^H(\rho) = \left(\sqrt{\rho}, T_{op}\sqrt{\rho} \right) - \frac{\kappa}{2} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x-y|^{-1} dx dy$$

for nonnegative ρ with $\int \rho = N$. This differs from the Thomas-Fermi functional in the replacement of $\int j(\rho)$ by (8.16). As usual, we define the **Hartree energy** to be

$$(8.19) \quad E^H(N) = \inf \left\{ \mathcal{E}^H(\rho) : \int \rho = N \right\}.$$

The scaling $\rho(x) \rightarrow N\rho(x)$ reduced the problem to one in which $\int \rho = 1$, $\kappa \rightarrow \kappa N \equiv \omega$ and $E \rightarrow NE$. From this we learn that the parameter $\omega = \kappa N$ is the crucial one for bosons, not $\tau = \kappa N^{2/3}$.

The minimization of (8.18) leads now to a genuine differential equation for $f = \sqrt{\rho}$

$$(8.20) \quad \{T_{op} - \omega|x|^{-1} * f^2\}f = -\mu f$$

on all of $\mathbf{R}^3$ and with μ chosen so that $\int f^2 = N$. It is proved in [LY2] that again there is a critical value ω_c such that when $\omega < \omega_c$ a minimizer exists and it is spherically symmetric and decreasing (although not with compact support). If $\omega > \omega_c$ then $E^H(N) = -\infty$. If we fix ω and let $\kappa \rightarrow \infty$ and $N \rightarrow \infty$ with $\omega = \kappa N$ then, as in the previous fermionic case, the solution to the quantum problem converges to the solution to the Hartree problem, i.e., $E(N)/E^H(N) \rightarrow 1$ and $\frac{1}{N}\rho_\psi - \frac{1}{N}\rho^H \rightarrow 0$.

Thus, *bosonic stars can collapse as well*, but whereas in the fermionic case the critical number N_c is proportional to $\kappa^{-3/2} \approx 10^{57}$, for bosons N_c is proportional to $\kappa^{-1} \approx 10^{38}$. The former defines an object which is the size of a *star*, while the latter defines only an object as massive as a *mountain* (assuming that the same value is used for the mass of the constituent particles.)

It is possible to understand this κ^{-1} behavior in a simple way—which also explains the central difficulty in proving Theorem 9. In (7.14) we saw that the kinetic energy of a particle ceases to control a $-C/|x|$ singularity whenever $C > 2/\pi$. In our bosonic star case we have many particles with the tiny constant $C \equiv \kappa$ between each pair. Suppose now that $(N-1)\kappa$ exceeds $2/\pi$ and that $N-1$ particles come together at a common point. Then the N^{th} particle feels an over-critical attraction and “falls in the hole.” But then every particle is trapped and none can escape. Something like this artificial scenario is what happens in the bosonic case and leads to collapse if $\omega = N\kappa$ is big enough. (It is not claimed that $\omega_c = 2/\pi$; in fact it is known only that $4/\pi < \omega_c < 2.7$.) The interesting fact about fermions is that the Pauli exclusion principle prevents this scenario from happening. Since $N \approx \kappa^{3/2}$ in this case, it would require the coalescence of only a tiny fraction of the particles (namely $N^{2/3}/N$) in order to form a “trap” for the remaining particles—but such a “fluctuation” does not occur with any significant probability.

Now let us return to the second kind of cold star—the **white dwarf**. In 1914 Adams discovered that while the companion star of Sirius has a mass about equal to that of the sun (which was known from its perturbation of Sirius) its radius had to be very small [ST]. He inferred this from the fact that the star is hot (in fact its light is whiter than that of most stars—hence the name) but it has a very small total luminosity. Many such stars are now known and they are thought to be burned out stars consisting of ordinary electrons and nuclei such as we discussed in Part VII. Their radius is bigger than that of a neutron star—about 5000 km—for a reason that will soon be apparent, and this means that general relativistic effects are not important here. The ones we can see optically shine because they have not yet rid themselves of all the energy of gravitational collapse.

For some time it was a real puzzle to explain why this dead star did not collapse entirely. We have seen that the quantum mechanical kinetic energy provides the requisite “internal pressure”, and the person who modelled this correctly was Chandrasekhar in

1931 [C]. There is, however, a slight twist from the neutron star model, which requires some explanation.

The main force among electrons and nuclei is electrostatic—not gravitational. But with a grain of faith we can suppose that the electric potential is cancelled locally, i.e., local neutrality is strongly enforced. This leaves only the gravitational potential, which is additive because there are no $\pm$ signs. Ideally we would like to incorporate both forces, but we shall retain only the gravitational force. The nuclei, as before, can be considered to be almost motionless but the electrons move with high speeds (because of the high density) and therefore have to be treated relativistically.

With this information we can, with another grain of faith, suppose that the Hamiltonian H_N in (8.2) is relevant, but with the following identifications. The number of electrons is N and the mass m that appears in the kinetic energy, T_{op} , is the *electron* mass. On the other hand, the gravitational force comes mostly from the nuclei (because of their large mass), and therefore $\kappa = GM^2$ with M now being the mass in the star per electron, i.e., $M = M_{nucleus}/z_{nucleus}$, which is much bigger than the electron's mass and which is typically about twice M_n . It is the presence of the very small m in the kinetic energy which make the radius bigger for the solution of (8.5), (8.6); by scaling, the radius is proportional to $1/m$ for a fixed N . The critical τ does not change (except insofar as $M \neq M_n$).

The proper model to analyze, of course, would be one with two kinds of particles (electrons and nuclei) and with two kinds of forces (electrostatic and gravitational), but this remains an open problem. There are some remarks about this problem in [LT3]. Another thing one would like to analyze is the effect of positive temperature. For *nonrelativistic* particles, both kinds of modifications have been achieved, in a certain limit, by Hertel, Narnhofer and Thirring [HT, HNT, TW]. See also Messer's book [M]. The relativistic case with both kinds of forces is still an open problem.

Chandrasekhar, of course, did not analyze the Hamiltonian H_N as we did here. He started instead with the Euler-Lagrange equation of the Thomas-Fermi approximation, (8.14) or (8.15). In fact, it was the following physical interpretation of (8.14) that motivated Chandrasekhar: If we take the gradient of both sides and then multiply by $-\rho(x)$, the right side becomes the gravitational force at a point x in the star. The left side can be interpreted as the gradient of the pressure, P , of a "quantum mechanical fluid." This fluid has an "equation of state" (i.e., the pressure, P as a function of the density, ρ) given by the following formula that is

valid for all fluids.

$$(8.21) \quad P(\rho) = \rho j'(\rho) - j(\rho) = \rho^2 \frac{\partial}{\partial \rho} \frac{j(\rho)}{\rho}.$$

This balancing of forces, i.e., the “gravitational-hydrostatic equilibrium,” was Chandrasekhar’s starting point.

From this point of view there is an interesting contrast between the bosonic case (well approximated by the Hartree theory (8.18)) and the fermionic case (well approximated by the Thomas-Fermi theory (8.5)). The latter energy can be thought of as that of a simple fluid which has an “equation of state.” This is how physicists think of the matter, even though it is quantum mechanics that produces the pressure. The bosonic energy (8.18) has no such interpretation. Quantum mechanics shows itself to the bitter end because the Hartree energy still has gradients in it. The concept of an “equation of state” cannot be used for high density bosonic matter in its ground state.

This brings us to the end of the stability of matter story. If one more hyperbolic remark be permitted, it can be said that

Quantum mechanics is a bizarre theory, invented to explain atoms. As far as we know today it is capable of explaining everything about ordinary matter (chemistry, biology, superconductivity), sometimes with stunning numerical accuracy. But it also says something about the occurrence of the most spectacular event in the cosmos—the supernova. The range is 57 orders of magnitude!

REFERENCES

- [AI] *American Institute of Physics Handbook*, McGraw-Hill, New York, 1972 third ed., p. 7–6.
- [AM] P. Armbruster and G. Münzenberg, *Creating superheavy elements*, Scientific American **260** (1989), 66–72.
- [AB] J. Auchmuty and R. Beals, *Variational solution of some nonlinear free boundary problems*, Arch. Rat. Mech. Anal. **43** (1971), 255–271. See also *Models of rotating stars*, Astrophys. J. **165** (1971), L79–L82.
- [B] G. Baym, *Neutron stars*, in Encyclopedia of Physics, (R. G. Lerner and G. L. Trigg eds.) Addison-Wesley, London, 1981, pp. 659–660.
- [BM] M. Born, *Quantenmechanik der Stossvorgänge*, Z. Phys. **38** (1926), 803–827.
- [CH] S. Chandrasekhar, *The maximum mass of ideal white dwarfs*, Astrophys. J. **74** (1931), 81–82. See also *On stars, their evolution and stability*, Rev. Mod. Phys. **56** (1984), 137–147.
- [CO] J. Conlon, *The ground state energy of a classical gas*, Comm. Math. Phys. **94** (1984), 439–458.
- [CLY] J. G. Conlon, E. H. Lieb and H-T. Yau, *The $N^{7/5}$ law for charged bosons*, Comm. Math. Phys. **116** (1988), 417–448.

- [DA] I. Daubechies, *An uncertainty principle for fermions with generalized kinetic energy*, Comm. Math. Phys. **90** (1983), 511–520.
- [DAL] I. Daubechies and E. H. Lieb, *One electron relativistic molecules with Coulomb interactions*, Comm. Math. Phys. **90** (1983), 497–510.
- [D] F. J. Dyson, *Ground state energy of a finite system of charged particles*, J. Math. Phys. **8** (1967), 1538–1545.
- [DL] F. J. Dyson and A. Lenard, *Stability of matter. I and II*, J. Math. Phys. **8** (1967), 423–434; *ibid* **9** (1968), 698–711.
- [FD] C. Fefferman and R. de la Llave, *Relativistic stability of matter. I.*, Rev. Math. Iberoamericana **2** (1986), 119–215.
- [FE] E. Fermi, *Un metodo statistico per la determinazione di alcune priorieta dell'atomo*, Atti Acad. Naz. Lincei, Rend. **6** (1927), 602–607.
- [FR] A. P. French, *Atoms*, in Encyclopedia of Physics, (R. G. Lerner and G. L. Trigg eds.), Addison-Wesley, London (1981), p. 64.
- [H] I. Herbst, *Spectral theory of the operator $(p^2 + m^2)^{1/2} - ze^2/r$* , Comm. Math. Phys. **53** (1977), 285–294. Errata, *ibid* **55** (1977), 316.
- [HNT] P. Hertel, H. Narnhofer and W. Thirring, *Thermodynamic functions for fermions with gravostatic and electrostatic interactions*, Comm. Math. Phys. **28** (1972), 159–176.
- [HT] P. Hertel and W. Thirring, *Free energy of gravitating fermions*, Comm. Math. Phys. **24** (1971), 22–36.
- [J] J. H. Jeans, *The mathematical theory of electricity and magnetism*, Cambridge Univ. Press, Cambridge, third edition, 1915, p. 168.
- [JM] M. Jammer, *The conceptual development of quantum mechanics*, McGraw-Hill, New York, 1966.
- [K] T. Kato, *Perturbation theory for linear operators*, Springer-Verlag, Heidelberg, 1966. See Remark 5.12 on p. 307.
- [KS] H. Kalf, U.-W. Schminke, J. Walter and R. Wüst, *On the spectral theory of Schrödinger and Dirac operators with strongly singular potentials*, Lecture Notes in Math., vol. **448** Springer-Verlag, Berlin and New York, 1974, pp. 182–226.
- [LE] A. Lenard, *Lectures on the Coulomb stability problem*, Lecture Notes in Physics **20** (1973), 114–135.
- [L1] E. H. Lieb, *Stability of matter*, Rev. Mod. Phys. **48** (1976), 553–569.
- [L2] —, *On characteristic exponents in turbulence*, Comm. Math. Phys. **92** (1984), 473–480.
- [L3] —, *Thomas-Fermi and related theories of atoms and molecules*, Rev. Mod. Phys. **53** (1981), 603–641; errata *ibid* **54** (1982), 311.
- [L4] —, *Bound on the maximum negative ionization of atoms and molecules*, Phys. Rev. **29A** (1984), 3018–3028. A summary is in Phys. Rev. Lett. **52** (1984), 315–317.
- [L5] —, *The $N^{5/3}$ law for bosons*, Phys. Lett. A **70** (1979), 71–73.
- [LL] E. H. Lieb and J. L. Lebowitz, *The constitution of matter: existence of thermodynamics for systems composed of electrons and nuclei*, Adv. in Math. **9** (1972), 316–398.
- [LO] E. H. Lieb and S. Oxford, *An improved lower bound on the indirect Coulomb energy*, Int. J. Quant. Chem. **19** (1981), 427–439.

- [LS] E. H. Lieb and B. Simon, *The Thomas-Fermi theory of atoms, molecules and solids*, Adv. in Math. **23** (1977), 22–116.
- [LT1] E. H. Lieb and W. E. Thirring, *Bound for the kinetic energy of fermions which proves the stability of matter*, Phys. Rev. Lett. **35** (1975), 687–689. Errata ibid. **35** (1975), 1116.
- [LT2] —, *Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities*, in *Studies in Mathematical Physics*, (E. Lieb, B. Simon and A. Wightman, eds.), Princeton Univ. Press, Princeton, New Jersey, 1976, pp. 269–330.
- [LT3] —, *Gravitational collapse in quantum mechanics with relativistic kinetic energy*, Ann. of Phys. (NY) **155** (1984), 494–512.
- [LY1] E. H. Lieb and H.-T. Yau, *The stability and instability of relativistic matter*, Comm. Math. Phys. **118** (1988), 177–213. A summary is in *Many-body stability implies a bound on the fine structure constant*, Phys. Rev. Lett. **61** (1988), 1695–1697.
- [LY2] —, *The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics*, Comm. Math. Phys. **112** (1987), 147–174. A summary is in *A rigorous examination of the Chandrasekhar theory of stellar collapse*, Astrophys. J. **323** (1987), 140–144.
- [M] J. Messer, *Temperature dependent Thomas-Fermi theory*, Lectures Notes in Physics no. 147, Springer-Verlag, Berlin and New York, 1981.
- [P] W. Pauli, *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*, Z. Phys. **31** (1925), 765–785.
- [RB] R. Ruffini and S. Bonazzola, *Systems of self-gravitating particles in general relativity and the concept of equation of state*, Phys. Rev. **187** (1969), 1767–1783.
- [SE] E. Schrödinger, *Quantisierung als Eigenwertproblem*, Ann. Phys. **79** (1926), 361–376. See also ibid. **79** (1926), 489–527; **80** (1926), 437–490; **81** (1926), 109–139.
- [ST] S. L. Shapiro and S. A. Teukolsky, *Black holes, white dwarfs and neutron stars*, Wiley, New York, 1983.
- [T] L. H. Thomas, *The calculation of atomic fields*, Proc. Cambridge Philos. Soc. **23** (1927), 542–548.
- [TW] W. Thirring, *A course in mathematical physics*, vol. 4, Springer-Verlag, Berlin and New York, 1983.

DEPARTMENTS OF MATHEMATICS AND PHYSICS, PRINCETON UNIVERSITY,
PRINCETON, NEW JERSEY 08544

Part II

Exact Results on Atoms

Lower bound to the energy of complex atoms

Peter Hertel

Institut für Theoretische Physik der Universität Wien, A-1090 Wien, Boltzmannngasse 5, Austria

Elliot H. Lieb*

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08540

Walter Thirring

Institut für Theoretische Physik der Universität Wien, A-1090 Wien, Boltzmannngasse 5, Austria

(Received 31 January 1975)

There are methods available for calculating rather precise lower bounds for the energy of simple atoms or molecules.¹ If complex atoms or molecules are to be investigated, these methods become inapplicable, or impracticable. Here we propose a new method which, for the above cases, promises to become as accurate as the Hartree-Fock procedure for the upper bound. Although applicable to a wider class of problems, we shall demonstrate it for an atom with N electrons.

Our aim is to find a lower bound to the Coulomb repulsion energy

$$V = \frac{1}{2} \sum_{\substack{a,b=1 \\ a \neq b}}^N \frac{1}{|\mathbf{x}_a - \mathbf{x}_b|} \quad (1)$$

of the form

$$V_{LB} = \sum_{a=1}^N \Phi(\mathbf{x}_a) - \Lambda, \quad (2)$$

where Λ is a constant. To achieve this we replace the Coulomb potential $1/|\mathbf{x}|$ by a subsidiary potential v with the following properties:

$$\frac{1}{|\mathbf{x}|} \geq v(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\mathbf{x}} \tilde{v}(\mathbf{k}) \quad \text{with} \quad \tilde{v}(\mathbf{k}) \geq 0. \quad (3)$$

$\tilde{v}$, the Fourier transform of v , is assumed to be integrable so that v is continuous everywhere. The repulsion potential Φ should be the Fourier transform of a real integrable function:

$$\Phi(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\mathbf{x}} \tilde{\Phi}(\mathbf{k}) \quad \text{with} \quad \tilde{\Phi}(\mathbf{k}) = \Phi(-\mathbf{k}) = \tilde{\Phi}^*(\mathbf{k}). \quad (4)$$

From (3) we deduce

$$V \geq \frac{1}{2} \sum_{a,b=1}^N v(\mathbf{x}_a - \mathbf{x}_b) - \frac{1}{2} N v(0). \quad (5)$$

With Jensen's or Schwarz's inequality

$$\int d\mu(x) |f(x)|^2 \geq \frac{|\int d\mu(x) f(x)|^2}{\int d\mu(x)} \quad (6)$$

(μ a measure) we deduce with (3)

$$\begin{aligned} \sum_{a,b=1}^N v(\mathbf{x}_a - \mathbf{x}_b) &= \int \frac{d^3k}{(2\pi)^3} \tilde{v}(\mathbf{k}) \left| \sum_{a=1}^N e^{i\mathbf{k}\mathbf{x}_a} \right|^2 \\ &= \int \frac{d^3k}{(2\pi)^3} \frac{\tilde{\Phi}(\mathbf{k})^2}{\tilde{v}(\mathbf{k})} \left| \sum_{a=1}^N e^{i\mathbf{k}\mathbf{x}_a} \frac{\tilde{v}(\mathbf{k})}{\tilde{\Phi}(\mathbf{k})} \right|^2 \geq \frac{1}{c} U^2, \end{aligned} \quad (7)$$

where

$$U = \sum_{a=1}^N \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\mathbf{x}_a} \tilde{\Phi}(\mathbf{k}) = \sum_{a=1}^N \Phi(\mathbf{x}_a), \quad (8)$$

$$c = \int \frac{d^3k}{(2\pi)^3} \frac{\tilde{\Phi}(\mathbf{k})^2}{\tilde{v}(\mathbf{k})} \quad (9)$$

Note that U is Hermitian and c real.

By subtracting the nonnegative operator $(1/c)(U - c)^2$ from the rhs of (7) we finally arrive at the desired bound:

$$V \geq \sum_{a=1}^N \Phi(\mathbf{x}_a) - \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} \left\{ \frac{\tilde{\Phi}(\mathbf{k})^2}{\tilde{v}(\mathbf{k})} + N v(\mathbf{k}) \right\}. \quad (10)$$

There are two functions to play with: a repulsion potential Φ (essentially arbitrary) and a subsidiary potential v [restricted by (3)] which enters the constant only.

A special choice for the subsidiary potential is

$$v(\mathbf{x}) = \frac{1 - e^{-\mu|\mathbf{x}|}}{|\mathbf{x}|}. \quad (11)$$

Defining n by

$$\Phi(\mathbf{x}) = \int d^3x' \frac{n(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \quad (12)$$

we can write

$$\Lambda = \frac{1}{2} \left\{ \int d^3x \Phi(\mathbf{x}) n(\mathbf{x}) + 4\pi\mu^{-2} \int d^3x n(\mathbf{x})^2 + N\mu \right\} \quad (13)$$

for the constant in the rhs of (10). For fixed n (or Φ) the optimal subsidiary potential of type (11) is obtained for

$$\mu^3 = \frac{8\pi}{N} \int d^3x n(\mathbf{x})^2. \quad (14)$$

To check whether (10) can give good lower bounds we have tried the Tietz approximation² to the Thomas-Fermi potential for neutral atoms:

$$\Phi_r(\mathbf{x}) = (N-1) \frac{r + 2(9/2N)^{1/3}}{[r + (9/2N)^{1/3}]^2} \quad (r = |\mathbf{x}|). \quad (15)$$

The factor $N-1$ (instead of N) in (15) has been chosen in order to make the potential of the lower bound Hamiltonian to go as $-1/r$ for large r . Equation (15) with (14) gives

$$\Lambda = \frac{2}{5} \left(\frac{2}{9} \right)^{1/3} \left(1 - \frac{1}{N} \right)^{2/3} N^{7/3} + \frac{3}{2} \left(\frac{2}{9} \right)^{1/3} \left(1 - \frac{1}{N} \right)^{2/3} N^{5/3}. \quad (16)$$

It remains to solve for the energy levels of the one-particle

Schrödinger equation with spherically symmetric potential $-N/|\mathbf{x}| + \Phi_T(\mathbf{x})$ and to fill the N electrons into the lowest levels in accordance with the exclusion principle. The lower bound E_{LB} for the nonrelativistic ground state energy of neutral atoms as calculated from (10) with (11), (14), and (15) is compared with an upper bound³ E_{UB} for a few neutral atoms in the following table:

Atom	N	$-N^{-7/3} E_{LB}$	$-N^{-7/3} E_{UB}$
Ne	10	0.755	0.594
Ca	20	0.726	0.620
Zr	40	0.714	0.646
Nd	60	0.711	0.648
Hg	80	0.711	0.656

Energies are in units of $me^4/\hbar^2 = 27.21$ eV. The uncorrected Thomas-Fermi value (which becomes exact⁴ in the limit $N \rightarrow \infty$) is $N^{-7/3} E_{TF} = -0.769$. The trivial lower bound (by neglecting V altogether) is $-(3/2)^{1/3} N^{1/3}$ for

large N . Note that the lower and upper bounds are still far away from the Thomas-Fermi value.

The increase with N of the lower bounds is due to the fact that the potential (15) is tailored for large atoms. By adjusting some parameters the lower bounds for light atoms can be improved so that they decrease with N .

*On leave from the Departments of Mathematics and Physics, M.I.T., Cambridge, MA 02139. Work partially supported by National Science Foundation Grant GP-31674X.

¹For a summary see A. Weinstein and W. Stenger, *Methods of Intermediate Problems for Eigenvalues* (Academic, New York, 1972).

²T. Tietz, J. Chem. Phys. **25**, 787 (1956). See S. Flügge, *Practical Quantum Mechanics II* (Springer, Berlin, 1971), p. 128, for the choice of the coefficient $(9/2)^{1/3}$.

³From R. O. Mueller, A. R. P. Rau, and L. Spruch, Phys. Rev. A **8**, 1186 (1973), Table II.

⁴E. H. Lieb and B. Simon, Phys. Rev. Lett. **31**, 681 (1973).

Improved Lower Bound on the Indirect Coulomb Energy*

ELLIOTT H. LIEB AND STEPHEN OXFORD

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08544, U.S.A.

Abstract

For a Coulomb system of particles of charge e , it has previously been shown that the indirect part of the repulsive Coulomb energy (exchange plus correlation energy) has a lower bound of the form $-Ce^{2/3} \int \rho(x)^{4/3} dx$, where ρ is the single particle charge density. Here we lower the constant C from the 8.52 previously given to 1.68. We also show that the best possible C is greater than 1.23.

1. Introduction

In the study of quantum Coulomb systems of charged particles (atoms, molecules, and solids), it is frequently desirable to estimate various energies in terms of the (diagonal) single particle charge density $\rho_\psi(x)$ belonging to a given state ψ of N particles. We will be concerned with the repulsive Coulomb energy

$$I_\psi = \langle \psi | \sum_{i < j}^N e_i e_j |x_i - x_j|^{-1} | \psi \rangle, \quad (1)$$

where $e_i > 0$ are the particle charges (by a trivial change we could assume all $e_i < 0$). ψ is *any* normalized state, not necessarily an eigenstate of any given Hamiltonian; it could also be a density matrix. Our goal is to find a lower bound to I_ψ . [See note added in proof.]

We define

$$D_\psi = D(\rho_\psi, \rho_\psi) \quad (2)$$

to be the *direct part* of I_ψ , where, in general, we have

$$D(f, g) = \frac{1}{2} \iint f(x)g(y)|x - y|^{-1} dx dy. \quad (3)$$

The density ρ_ψ is the sum $\sum_{i=1}^N$ of the N individual single particle charge densities defined by

$$\rho_\psi^i(x) = e_i \sum_{\alpha_1, \dots, \alpha_N} \int |\psi(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_N, \alpha_1, \dots, \alpha_N)|^2 d\hat{x}_i. \quad (4)$$

* Work partially supported by U.S. National Science Foundation Grant No. PHY-7825390 A01.

Here the α_j are any quantum numbers (e.g., spin) and $d\hat{x}_i$ means integration over all variables except x_i . The *indirect part* E_ψ of I_ψ is defined by

$$E_\psi + D_\psi = I_\psi. \quad (5)$$

E_ψ includes all "exchange" and "correlation" effects.

It was shown by one of us [1], that for all ψ and $C = 8.52$, we have

$$E_\psi \geq -Ce^{2/3} \int \rho_\psi(x)^{4/3} dx. \quad (6)$$

A more complicated bound was given when the e_i are not identical.

The main result of this paper is given in Section 2. We show that

$$E_\psi \geq -C \left[\int \left(\sum_{i=1}^N e_i \rho_\psi^i(x) \right)^{4/3} dx \right]^{1/2} \left[\int \rho_\psi(x)^{4/3} dx \right]^{1/2}, \quad (7)$$

even in the case that the e_i may be distinct, with $C = 1.68$. With this improved value of C , the bound for E_ψ may have computational value as well as theoretical value.

Inequality (7) reduces to the form of Eq. (6) when all $e_i = e$. Our proof differs from that in Ref. 1 chiefly in avoiding the use of the Hardy–Littlewood maximal function. We also show in Section 3 that the best (i.e., smallest) possible C , in Eq. (6), must be greater than 1.23 by explicitly calculating E for a two-particle wave function ψ .

It should be noted that Eq. (6) is a special case of the lower bound for N -particle states ψ ,

$$E_\psi \geq -C_p e^{(5p-6)/(3p-3)} \left(\int \rho_\psi^p \right)^{1/(3p-3)} N^{(3p-4)/(3p-3)}, \quad (8)$$

which is a consequence of Thomas–Fermi theory for $p > \frac{3}{2}$ (see Refs. 1–3). Hölder's inequality applied to Eq. (6) gives Eq. (8) for all $p \geq \frac{4}{3}$ with the same C as in Eq. (6). However, the best C_p in Eq. (8) will in general be smaller than this C and it would be interesting to have an estimate for it. [In Ref. 1 it was incorrectly stated that Thomas–Fermi theory gives Eq. (8) for $p > \frac{4}{3}$. This is wrong because Thomas–Fermi theory is bounded below only for $p > \frac{3}{2}$. Equation (8), for $p = \frac{5}{3}$, was first given in Ref. 3, where it was used to prove the stability of matter.]

The bound, equations (6) or (7), can be examined in a number of ways. One might try to find a constant C in Eq. (6) only for ψ symmetric or antisymmetric, or only for ψ with a particular spin value. However any such restriction on ψ cannot improve the constant in Eq. (6). Let all $e_i = e$. Suppose we take an arbitrary $\psi(x_1, \dots, x_N; \alpha_1, \dots, \alpha_N)$, where $\alpha_1, \dots, \alpha_N$ are arbitrary quantum numbers. Let

$$f_\psi(x_1, \dots, x_N) = \sum_{\alpha_1 \dots \alpha_N} |\psi(x_1, \dots, x_N, \alpha_1, \dots, \alpha_N)|^2,$$

and let the symmetrized version of f_ψ be given as

$$F(x_1, \dots, x_N) = \frac{1}{N!} \sum_{P \in S_N} f_\psi(x_{P1}, \dots, x_{PN}).$$

We can define a symmetric ψ^S such that $E_{\psi^S} = E_\psi$, $I_{\psi^S} = I_\psi$, and $\rho_{\psi^S}(x) = \rho_\psi(x)$. We merely take $\psi^S(x_1, \dots, x_N) = F(x_1, \dots, x_N)^{1/2}$. This shows that C cannot be improved by excluding bosonic ψ . In a similar way, we define an antisymmetric ψ^a ,

$$\psi^a = F(x_1, \dots, x_N)^{1/2} \theta(x_1, \dots, x_N),$$

where θ is any antisymmetric function which takes on the values ± 1 except on a set of measure zero. We see that we cannot improve the constant by excluding ψ with Fermi statistics. A similar construction shows that restrictions on the spin quantum number cannot improve C .

It will be noted that the right-hand side of Eq. (6) is of the form given by Dirac [4] to approximate the exchange energy. There is, however, a difference between the two if spin is taken into account. The Dirac approximation is

$$E_\psi \approx -Cq^{-1/3} \int \rho_\psi(x)^{4/3} dx, \quad (9)$$

with $C = 0.93$ and q is the number of spin states ($q = 2$ for electrons). Dirac computed Eq. (9) using a plane wave determinant for ψ . This determinant depends upon q . In view of Eq. (9) one might infer that a good lower bound to E_ψ should have a $q^{-1/3}$ factor. This, as we have just noted above, is not the case. Another way to say this is the following: a “diagonal” operator, such as the Coulomb repulsion cannot distinguish the spin of a particle. To “see” the spin it is necessary to examine an off-diagonal operator such as the kinetic energy T_ψ . A lower bound for T_ψ does exist [3] and it does have a factor $q^{-2/3}$. In order to have a lower bound for E_ψ that measures q it would be necessary to use an expectation value such as T_ψ that is “off-diagonal.” A useful bound of this kind might exist, but a bound such as ours that involves only the “diagonal” quantity ρ_ψ has no q dependence.

It is true however, that the constant in Eq. (6) or (7) can be improved if we specify the number of particles. Consider the case of equal charges. Define C_N to be the best constant for Eq. (6) when we consider only N -particle states ψ , i.e.,

$$C_N = \sup_{N \text{ particle } \psi} \left(- \int \rho_\psi(x)^{4/3} dx \right)^{-1} E_\psi.$$

In Appendix A we compute $C_1 = 1.092$, following a treatment by Gadre, Bartolotti, and Handy [5]. One notes that C_1 is less than the constant 1.23 computed for the variational two-particle wave function of Section 3, which is itself less than C_2 . In Appendix B, we show that this is a general phenomenon, i.e., $C_N \leq C_{N+1}$. Presumably, one might take advantage of this if one were only dealing with a fixed low number of particles. However, we have no method available for computing an upper bound for C_N , $N \geq 2$, except 1.68, which holds for all N .

Inequality (7) might be improved upon in another way in the case of unequal charges. One might try to prove that a universal constant C existed such that the following is true (for fixed $0 < \alpha \leq 1$):

$$E_\psi \geq -C \left[\int \left(\sum_{i=1}^N e_i^{1/2\alpha} \rho_\psi^i(x) \right)^{4/3} dx \right]^\alpha \left[\int \rho(x)^{4/3} dx \right]^{1-\alpha}. \quad (10)$$

It is an open question whether a bound exists when $\alpha = 1$ in Eq. (10), called the symmetric form. Hölder's inequality implies that for Eq. (10), (right-hand side, $\alpha = 1$) $\geq$ (right-hand side, $\alpha \leq 1$); thus the symmetric form gives the best lower bound. In Ref. 1 a bound with constant 8.52 was proved for $\alpha = \frac{3}{4}$ and in this paper we sacrifice even more "symmetry" in the bound to show that Eq. (10) holds for $\alpha = \frac{1}{2}$ with the improved constant 1.68.

In Ref. 5, the comment was made that no one has yet produced any upper bound for E_ψ . The following simple remark is relevant. There certainly cannot be any upper bound of form $C \int \rho(x)^{4/3} dx$. The reason is that $E_\psi / \int \rho(x)^{4/3} dx$ can be made arbitrarily positive simply by taking a ψ for two particles of the form

$$\psi = A[k \cdot (x_1 - x_2)] \exp(-|x_1|^2 - |x_2|^2 - \lambda |x_1 - x_2|^2),$$

where k is some fixed vector. As λ tends to infinity, E_ψ will tend to $+\infty$ while $\int \rho_\psi^{4/3} dx$ will remain bounded. This ψ is antisymmetric.

2. A Lower Bound for E_ψ

We will use an argument similar to that given in Ref. 1 to derive Eq. (7) for $C = 1.68$.

We first fix charges $e_1 \geq 0, \dots, e_N \geq 0$. Let $f_\psi(x_1, \dots, x_N)$ be the particle density associated with an N -particle wave function ψ , as given in Section 1. Let ρ_ψ be the associated single-particle charge density, equation (4), and let $x_1, \dots, x_N$ be distinct but otherwise arbitrary points in $\mathbf{R}^3$. We take $\mu(y)$ to be some function satisfying the following: (i) μ is non-negative; (ii) μ is spherically symmetric about the origin and $\mu(x) = 0$ if $|x| > 1$; and (iii) $\int \mu(y) dy = 1$. Let λ be a positive constant to be determined later.

We now define a function

$$\mu_x(\cdot): \quad \mu_x(y) \equiv \lambda^3 \rho_\psi(x) \mu(\lambda \rho_\psi(x)^{1/3}(y - x)), \quad (11)$$

if $\rho_\psi(x) > 0$, and $\mu_x(y) \equiv 0$ if $\rho_\psi(x) = 0$. We see that μ_x is a non-negative function which satisfies (i) its integral (with respect to y) is 1 if $\rho_\psi(x) > 0$; (ii) it is spherically symmetric about x ; (iii) $\mu_x(y) = 0$ if $|y - x| > \lambda^{-1} \rho_\psi(x)^{-1/3}$.

We observe that Lemma 1 of Ref. 1 may be applied to this choice of μ . Namely, we prove

Lemma 1:

$$\sum_{1 \leq i < j \leq N} e_i e_j |x_i - x_j|^{-1} \geq -D(\rho_\psi, \rho_\psi) + 2 \sum_{i=1}^N D(\rho_\psi, e_i \mu_{x_i}) - \sum_{i=1}^N D(e_i \mu_{x_i}, e_i \mu_{x_i}). \quad (12)$$

Proof. It is a well-known fact [1] that the potential ϕ generated by a non-negative spherically symmetric charge distribution of total charge 1 satisfies $\phi(x) \leq 1/|x|$. In particular, taking x_i to be the origin, $\int \mu_{x_i}(y)|x-y|^{-1} dy \leq |x-x_i|^{-1}$. Hence

$$\begin{aligned} D(e_i \mu_{x_i}, e_j \mu_{x_j}) &= \frac{1}{2} e_i e_j \int \mu_{x_i}(y) |x-y|^{-1} \mu_{x_j}(x) dx dy \\ &\leq \frac{1}{2} e_i e_j \int |x-x_i| \mu_{x_j}(x) dx \leq \frac{1}{2} e_i e_j |x_i-x_j|^{-1}. \end{aligned} \quad (13)$$

We now observe that

$$D\left(\rho_\psi - \sum_{i=1}^N e_i \mu_{x_i}, \rho_\psi - \sum_{i=1}^N e_i \mu_{x_i}\right) \geq 0 \quad (14)$$

by the positive definiteness of the Coulomb kernel. Expanding the left-hand side of Eq. (14) and rearranging, one has that

$$2 \sum_{i < j} D(e_i \mu_{x_i}, e_j \mu_{x_j}) \geq -D(\rho_\psi, \rho_\psi) + 2 \sum_{i=1}^N D(\rho_\psi, e_i \mu_{x_i}) - \sum_{i=1}^N D(e_i \mu_{x_i}, e_i \mu_{x_i}). \quad (15)$$

The lemma follows by applying Eq. (13) to the left-hand side of Eq. (15). ■

Now let δ_{x_i} be a point charge distribution of charge one centered at x_i . Adding and subtracting terms on the right-hand side of inequality (12), we get

$$\begin{aligned} \sum_{i < j} e_i e_j |x_i - x_j|^{-1} &\geq -D(\rho_\psi, \rho_\psi) + 2 \sum_{i=1}^N D(\rho_\psi, e_i \delta_{x_i}) \\ &\quad - \left(2 \sum_{i=1}^N D(\rho_\psi, e_i \delta_{x_i} - e_i \mu_{x_i}) + \sum_{i=1}^N D(e_i \mu_{x_i}, e_i \mu_{x_i}) \right). \end{aligned} \quad (16)$$

We now integrate Eq. (16) against $f_\psi(x_1, \dots, x_N)$, whence

$$\begin{aligned} I_\psi &\geq D(\rho_\psi, \rho_\psi) - \left(\sum_{i=1}^N \int 2D(\rho_\psi, \delta_{x_i} - \mu_{x_i}) \rho_\psi^i(x_i) dx_i \right. \\ &\quad \left. + \sum_{i=1}^N \int D(\mu_{x_i}, \mu_{x_i}) e_i \rho_\psi^i(x_i) dx_i \right). \end{aligned} \quad (17)$$

We wish to find an upper bound of the appropriate kind to the expression in large parentheses in Eq. (17). Let us denote the first sum in large parentheses as (*) and the second as (**). We rewrite (*) as follows:

$$(*) = \int \rho_\psi(x) 2D(\rho_\psi, \delta_x - \mu_x) dx = \iint \rho_\psi(y) F_\lambda(\rho_\psi(x), |x-y|) dx dy, \quad (18)$$

where we define

$$F_\lambda(a, r) = [ar^{-1} - \lambda a^{4/3} \phi(\lambda a^{1/3} r)], \quad (19)$$

and ϕ is the potential generated by our fixed μ . Hereafter we require that μ be bounded. In this case, $F_\lambda(a, r)$, considered as a function of a on $(0, \infty)$, satisfies the

following for $r > 0$; (i) it is continuously differentiable; (ii) $F_\lambda(0, r) = 0$ and $F_\lambda(a, r) = 0$ if $\lambda a^{1/3} r > 1$. These properties follow from the continuity and differentiability of ϕ and the relation $\phi(x) = 1/|x|$ for $|x| > 1$. For $a > 0$, let $\chi_a(x) = \theta[\rho_\psi(x) - a]$, where $\theta(t) = 0$ if $t \leq 0$, and $\theta(t) = 1$ if $t > 0$. χ_a is the characteristic function of the set $\rho_\psi(x) > a$. By Fubini's theorem and the fundamental theorem of calculus, one has that

$$\int_0^\infty da \int \chi_a(x) \frac{\partial}{\partial a} F_\lambda(a, r) dx = \int dx \int_0^{\rho(x)} \frac{\partial}{\partial a} F_\lambda(a, r) da = \int dx F_\lambda(\rho(x), r), \quad (20)$$

and thus

$$(*) = \int_0^\infty \int_0^\infty da db \int dx dy \chi_a(x) \chi_b(y) \frac{\partial}{\partial a} F_\lambda(a, |x - y|)$$

where we have used the representation $\rho(y) = \int_0^\infty db \chi_b(y)$.

We bound $(*)$ as follows: Let $(y)_+ = y$ if $y \geq 0$, and $(y)_+ = 0$ if $y \leq 0$. Then

$$\begin{aligned} (*) &\leq \int_0^\infty \int_0^\infty da db \int dx dy \chi_a(x) \chi_b(y) \left(\frac{\partial}{\partial a} F_\lambda(a, |x - y|) \right)_+ \\ &\leq \int_{a>b} da db \int dx dy \chi_a(x) \left(\frac{\partial}{\partial a} F_\lambda(a, |x - y|) \right)_+ \\ &\quad + \int_{b<a} da db \int dx dy \chi_b(y) \left(\frac{\partial}{\partial a} F_\lambda(a, |x - y|) \right)_+. \end{aligned} \quad (21)$$

By scaling properties of $(\partial/\partial a)F_\lambda(a, |x - y|)$, one has that

$$\int \left(\frac{\partial}{\partial a} F_\lambda(a, |x - y|) \right)_+ dx = \int \left(\frac{\partial}{\partial a} F_\lambda(a, |x - y|) \right)_+ dy = \lambda^{-2} K a^{-2/3}, \quad (22)$$

where $K = \int [(\partial/\partial a)F_1(1, |z|)]_+ dz$ and K only depends on the original choice of μ .

We, therefore, have that

$$\begin{aligned} (*) &\leq \lambda^{-2} K \int dx \left(\int_0^\infty da \chi_a(x) a^{-2/3} \int_0^a db + \int_0^\infty \chi_b(x) db \int_0^b a^{-2/3} da \right) \\ &= 4\lambda^{-2} K \int dx \int_0^\infty \chi_a(x) a^{1/3} da = 3\lambda^{-2} K \int \rho_\psi^{4/3}(x) dx, \end{aligned} \quad (23)$$

where we have used the representation $\rho_\psi^{4/3}(x) = (4/3) \int_0^\infty a^{1/3} \chi_a(x) da$.

The second sum $(**)$ in the large parentheses of (17) can be written

$$\begin{aligned} (**) &= \sum_{i=1}^N \int D(\mu_x, \mu_x) e_i \rho_\psi^i(x) dx \\ &= \lambda D(\mu, \mu) \sum_{i=1}^N \int e_i \rho_\psi^i(x) \rho_\psi(x)^{1/3} dx \end{aligned} \quad (24)$$

$$\leq \lambda D(\mu, \mu) \left[\int \left(\sum_{i=1}^N e_i \rho_\psi^i(x) \right)^{4/3} dx \right]^{3/4} \left[\int \rho_\psi(x)^{4/3} dx \right]^{1/4}. \quad (25)$$

Equation (24) follows from simple scaling and Eq. (25) is the Hölder inequality. Optimizing Eqs. (24) and (25) with respect to λ yields

$$(*) + (**) \leq \frac{3}{2} [6KD(\mu, \mu)^2]^{1/3} \left[\int \left(\sum_{i=1}^N e_i \rho^i(x) \right)^{4/3} dx \right]^{1/2} \left[\int \rho(x)^{4/3} dx \right]^{1/2}. \quad (26)$$

A variational argument shows that the optimum choice of μ would be the uniform ball if $[(\partial/\partial a)F_\lambda(a, r)]_+$ were replaced by $(\partial/\partial a)F_\lambda(a, r)$ [in which case the constant in Eq. (26) would be 1.45]. However, trial and error indicates this choice is also approximately best with the cutoff. We find that $[\partial F_1(1, r)/\partial a]_+ = \partial F_1(1, r)/\partial a$ if and only if $r \leq R$ with $R = (5^{1/2} - 1)/2$. Then $K = 0.6489$ and $D(\mu, \mu) = \frac{3}{5}$. The constant in Eq. (26) is then 1.68.

Thus we reach the conclusion that

$$E_\psi \geq -1.68 \left[\int \left(\sum_{i=1}^N e_i \rho_\psi^i(x) \right)^{4/3} dx \right]^{1/2} \left[\int \rho_\psi(x)^{4/3} dx \right]^{1/2}. \quad (27)$$

3. A Lower Bound for C_2

We now exhibit a lower bound to C_2 [and thus to the best possible C in Eq. (6)] which is greater than C_1 .

We choose a singular $\psi(x, y)$, and take $e_1 = e_2 = 1$. Let $t \doteq |x|$ and $s \doteq |y|$ for $x, y \in \mathbf{R}^3$, and let h and e be unit vectors $e = x/|x|$ and $h = y/|y|$. We define

$$|\psi|^2[(t, e), (s, h)] \equiv f \equiv (15/4\pi^2) \delta(1-t-s) \delta(e \cdot h + 1) \theta(1-t) \theta(1-s).$$

We check the following:

$$\begin{aligned} \int f[(t, e), (s, h)] s^2 ds dh &= \frac{15}{4\pi^2} \int_0^1 \delta(1-t-s) s^2 dx \\ &\quad \times \int_0^{2\pi} \int_0^\pi \delta(\cos \psi + 1) \sin \psi d\psi d\phi \theta(1-t) \\ &= \frac{15}{2\pi} (1-t)^2 \theta(1-t) \equiv \rho^1(t). \end{aligned}$$

We have used spherical coordinates to evaluate the above integral with the north pole in the (fixed) “ e ” direction. Similarly, the s marginal is $\rho^2(s) = \rho^1(s)$. One checks that $\int \rho^1 = 1$ and hence f is properly normalized. We have that $\rho(t) = 2\rho^1(t)$. Trivially, $I_\psi = 1$ since the particles are always one unit apart. We have that

$$\int \rho(x)^{4/3} dx = 4\pi \left(\frac{15}{\pi} \right)^{4/3} \int_0^1 (1-t)^{8/3} t^2 dt = 2.084.$$

By Newton’s theorem

$$D(\rho, \rho) = (4\pi)^2 \int_0^\infty \rho(t) t \int_0^t \rho(s) s^2 ds = 3.572.$$

The result is that

$$C \geq C_2 \geq -E_\psi / \int \rho_\psi(x)^{4/3} dx = 1.234.$$

Appendix A: Evaluation of C_1

For one particle $I_\psi = 0$. Therefore C_1 is the maximum of $H(\rho)$ given in Eq. (28). We prove the following theorem and then compute C_1 . This was done in Ref. 5. The purpose of this appendix is only to give a rigorous justification of the calculation done in Ref. 5.

Theorem. There exists a symmetric decreasing function ρ which maximizes the functional

$$H(\rho) = \left(\int \rho(x)^{4/3} dx \right)^{-1} D(\rho, \rho) \quad (28)$$

over the set $A = \{\rho(x) | \rho(x) \geq 0, \rho \in L^{4/3}(\mathbf{R}^3), \int \rho(x) dx = 1\}$.

Proof. For any $\rho \geq 0$ and $\rho \in L^{4/3}(\mathbf{R}^3) \cap L^1(\mathbf{R}^3)$, let $(\rho)_\lambda$ be a scaled version of ρ , i.e., $(\rho)_\lambda(x) = \lambda^3 \rho(\lambda x)$. It is simple to check that

$$\begin{aligned} \int (\rho)_\lambda(x) dx &= \int \rho(x) dx, \\ \int (\rho)_\lambda(x)^{4/3} dx &= \lambda \int \rho(x)^{4/3} dx, \end{aligned}$$

and

$$H[(\rho)_\lambda] = H(\rho). \quad (29)$$

Let $\rho_j \in A$, $j = 1, 2, \dots$ be such that $\lim_j H(\rho_j) = \sup_{\rho \in A} H(\rho)$. This supremum may *a priori* be infinite, but we will see that it is finite. By scaling ρ_j and using Eq. (29), we may assume henceforth that $\int \rho_j(x)^{4/3} dx = 1$. By the Riesz inequality for symmetric decreasing rearrangements [6],[†] we have $H(\rho^*) \geq H(\rho)$, where ρ^* is the symmetric decreasing rearrangement of ρ . One also has that $\int \rho^*(x) dx = \int \rho(x) dx$, $\int \rho^*(x)^{4/3} dx = \int \rho(x)^{4/3} dx$; therefore, by replacing ρ_j by ρ_j^* if necessary, we may also assume that the ρ_j are symmetrically decreasing.

We now use an idea used in Ref. 7. By the symmetric decreasing property of ρ_j , we have [writing $\rho_j(x) = \rho_j(|x|)$]

$$\frac{4\pi}{3} R^3 \rho_j(R) \leq \int_{|x| \leq R} \rho_j(x) dx \leq \int \rho_j(x) ds = 1.$$

[†] The 3-dimensional proof can be found in Brascamp, Lieb, and Luttinger, J. Funct. Anal. **17**, 227 (1974).

Hence

$$\rho_j(R) \leq k_1/R^3, \text{ all } j. \quad (30)$$

Similarly,

$$\begin{aligned} \frac{4\pi}{3} R^3 \rho_j^{4/3}(R) &\leq \int \rho_j(x)^{4/3} dx = 1, \\ \rho_j(R) &\leq k_2/R^{9/4}, \text{ all } j. \end{aligned} \quad (31)$$

We define $f(R) \equiv \min(k_1/R^3, k_2/R^{9/4})$. Since the ρ_j are symmetric decreasing and uniformly bounded by f (which is finite except at 0), by a variant of Helley's theorem [8], some subsequence of the ρ_j (which we continue to denote by ρ_j) converges pointwise almost everywhere to some symmetric decreasing $\rho(x)$ and $\rho(x) \leq f(x)$. We will see that $\rho(x) \neq 0$.

We now show that the ρ we have found satisfies the conditions of the theorem. By calculation $D(f, f) < \infty$. We therefore apply the dominated convergence theorem to conclude that

$$\lim_j D(\rho_j, \rho_j) = D(\rho, \rho) < \infty. \quad (32)$$

In particular,

$$0 < \sup_A H(\rho) = \lim D(\rho_j, \rho_j) < \infty \text{ and thus } \rho \neq 0. \quad (33)$$

Furthermore, by Fatou's lemma we have that

$$\int \rho(x)^{4/3} dx \leq \lim_j \int \rho_j(x)^{4/3} dx = 1, \quad (34)$$

$$\int \rho(x) dx \leq \lim_j \int \rho_j(x) dx = 1. \quad (35)$$

Therefore by Eqs. (32)–(34), $H(\rho) \geq \sup_{\rho \in A} H(\rho)$. By Eq. (35) we can multiply ρ by a scalar $\lambda \geq 1$ so that $\int \lambda \rho(x) dx = 1$. By definition of A , $\lambda \rho \in A$ and

$$H(\lambda \rho) = \lambda^{2/3} H(\rho) \geq H(\rho) \geq \sup_{\rho \in A} H(\rho). \quad (36)$$

It must be that the inequalities in Eq. (36) are actually equalities and thus that $\lambda = 1$. We therefore have that ρ belongs to A and maximizes $H(\cdot)$ on that set. ■

We shall now show that the constant C_1 can be calculated. By usual variational arguments [7], one knows that the optimizing ρ satisfies the following variational equations:

$$\phi(x) - \frac{4}{3} C_1 \rho(x)^{1/3} + \lambda \begin{cases} = 0 & \text{if } \rho(x) > 0, \\ \leq 0 & \text{if } \rho(x) = 0, \end{cases} \quad (37)$$

$$(38)$$

where $\phi(x)$ is the potential generated by ρ ,

$$\phi(x) = \int \rho(y)|x-y|^{-1} dy, \quad (39)$$

and where λ is a real-valued Lagrange multiplier.

We first note that Eqs. (37) and (38) imply that ρ is compactly supported. If not, then one must have that $\rho(x) > 0$ for all x by the symmetric decreasing property of ρ . By letting $|x| \rightarrow \infty$ one has that $\lambda = 0$, since ρ and ϕ tend to zero in Eq. (37). We then would have by Eq. (37), $\rho(x) = (\text{constant})\phi(x)^3$, where the constant is positive. For sufficiently large $|x|$, we see that $\rho(x) \geq \frac{1}{2}(\text{constant})|x|^{-3}$. This implies that ρ is not integrable, contradicting the fact that $\int \rho(x) dx = 1$. Let r_0 be the distance at which ρ first vanishes.

We now apply the Laplacian to Eq. (39) and use Eq. (37),

$$-(1/4\pi)\Delta\phi(x) = \rho(x) = \begin{cases} \{ \frac{3}{4}C_1^{-1}[\phi(x) + \lambda] \}^3, & |x| \leq r_0, \\ 0, & |x| \geq r_0. \end{cases} \quad (40)$$

Let $f(r) = (3\pi/C_1)^{3/2}[\phi(x) + \lambda]/4\pi$. We rewrite Eq. (40) in spherical coordinates

$$r^{-1} \frac{d^2}{dr^2} rf(r) = -f(r)^3, \quad r < r_0. \quad (41)$$

Equations (40) and (41) hold in the distributional sense.

We now argue that $f(r)$ is continuously differentiable and that $f'(0) = 0$. This will also imply that Eq. (41) is supplemented by $f(r_0) = 0$. In spherical coordinates, one can write

$$\phi(r) = \begin{cases} 4\pi r^{-1} \int_0^r \rho(s)s^2 ds + 4\pi \int_r^{r_0} \rho(s)s ds, & r \leq r_0, \\ r^{-1}, & r \geq r_0. \end{cases} \quad (42)$$

We apply Hölder's inequality to the first integral and use the fact that $\rho \in L^{4/3}$,

$$\int_0^r \rho(s)s^2 ds \leq \left(\int_0^r \rho(s)^{4/3}s^2 ds \right)^{3/4} \left(\int_0^r s^2 ds \right)^{1/4} \leq (\text{constant})r^{3/4}.$$

This and a similar inequality satisfied by the second integral imply that $\phi(r) = O(r^{-1/4})$ near the origin. By Eq. (37) one has that $\rho(r) = O(r^{-3/4})$ near $r = 0$. This in turn implies that ϕ is bounded at the origin, and therefore by Eq. (37), ρ is also bounded at the origin. Since ϕ is the potential of a bounded, compactly supported charge distribution, it is also continuous, hence ρ is continuous by Eqs. (37) and (38).

One now can see that ϕ (hence f) is C^1 by examining Eq. (42). Since ρ is continuous and bounded, the first term of Eq. (42) is of the form $r^{-1}g(r)$, where $g(r)$ is continuously differentiable for $r > 0$, $g(r) = O(r^3)$ and $g'(r) = O(r^2)$ near the origin. Hence the first term is continuously differentiable for $r \geq 0$, and has vanishing derivative at $r = 0$. The preceding statement is true of the second term in

Eq. (42) by inspection. Thus $\phi(r)$ is continuously differentiable for $r \geq 0$ and $\phi'(0) = 0$. Equation (41) holds in the strong sense because its right-hand side is C^1 .

As first noted by Gadre, Bartolotti, and Handy [5], Eq. (41) is the Emden equation of order 3. One may rescale $\rho(x) \rightarrow \alpha^3 \rho(\alpha x)$ to ensure that $f(0) = 1$. The two conditions $f(0) = 1$ and $f'(0) = 0$ uniquely determine the solution of the ordinary differential equation (41).

If r_0 is the first zero of the solution, we have that $\rho(r) = 0$ if $r \geq r_0$ and $\rho(r) = (3\pi)^{-3/2} C_1^{3/2} f(r)^3$ if $r \leq r_0$. In Ref. 5 it was noted that this equation determines the constant C_1 . Namely, we have that

$$\begin{aligned} 1 &= 4\pi \int_0^{r_0} \rho(r) r^2 dr = 4 \cdot 3^{-3/2} \pi^{-1/2} C_1^{3/2} \int_0^{r_0} r^2 f(r)^3 dr \\ &= -4 \cdot 3^{-3/2} \pi^{-1/2} C_1^{3/2} \int_0^{r_0} r [rf(r)]'' dr \\ &= -4 \cdot 3^{-3/2} \pi^{-1/2} C_1^{3/2} r_0^2 f'(r_0). \end{aligned} \quad (43)$$

Emden functions are tabulated [9]. We find that $r_0 = 6.89684$, $f'(r_0) = -0.04243$. Equation (43) then gives $C_1 = 1.092$.

Appendix B: Monotonicity of C_N

We show that $C_N \leq C_{N+1}$, where C_N is defined in Section 1 as the best constant in Eq. (6) for an N -particle state. We consider the case $e_i = e$.

Let $\varepsilon > 0$ be arbitrary but fixed. We let $f_N(x_1, \dots, x_N)$ be an N -particle density which vanishes for $|x_i| > L$ for $1 \leq i \leq N$, where L is some finite number, and furthermore, let f_N have the property that

$$-\left(e^{2/3} \int \rho_{f_N}(x)^{4/3} dx\right)^{-1} E_{f_N} \geq C_N - \varepsilon. \quad (44)$$

A simple approximation argument using dominated convergence shows that L and f_N can be found satisfying Eq. (44).

Let $x_0 \in \mathbf{R}^3$ be chosen such that $|x_0| \geq L + 2R$, where R will be determined later. We define a one-particle density $f_1(x) = (\frac{4}{3}\pi R^3)^{-1} \theta(R - |x - x_0|)$ and we also define the $(N+1)$ -particle density $f_{N+1}(x_1, \dots, x_{N+1}) = f_N(x_1, \dots, x_N) f_1(x_{N+1})$. One sees that $\rho_{f_{N+1}}(x) = \rho_{f_N}(x) + e f_1(x)$. Since ρ_N and f_1 are never simultaneously nonzero, we have

$$\begin{aligned} e^{2/3} \int \rho_{f_{N+1}}(x)^{4/3} dx &= e^{2/3} \left(\int \rho_{f_N}(x)^{4/3} dx + e^{4/3} \int f_1(x)^{4/3} dx \right) \\ &= e^{2/3} \int \rho_{f_N}(x)^{4/3} dx + \frac{e^2 (3/4\pi)^{1/3}}{R}. \end{aligned} \quad (45)$$

We also have that

$$\begin{aligned} I_{f_{N+1}} &= I_{f_N} + e^2 \sum_{i=1}^N \int f_N(x_1, \dots, x_N) f_1(x_{N+1}) |x_i - x_{N+1}|^{-1} dx_1, \dots, dx_{N+1} \\ &\leq I_{f_N} + e^2 N/R, \end{aligned} \quad (46)$$

by the definition of f_1 . The evident inequality $D_{f_{N+1}} \geq D_{f_N}$ together with Eq. (46) implies that $E_{f_{N+1}} \leq E_{f_N} + e^2 N/R$.

This and Eq. (45) imply that

$$\begin{aligned} C_{N+1} &\geq - \left(e^{2/3} \int \rho_{f_{N+1}}(x)^{4/3} dx \right)^{-1} E_{f_{N+1}} \\ &\geq - \left(e^{2/3} \int \rho_{f_N}(x)^{4/3} dx + \frac{e^2(3/4\pi)^{1/3}}{R} \right)^{-1} \left(E_{f_N} + \frac{e^2 N}{R} \right). \end{aligned} \quad (47)$$

We now choose R so large that the right-most term in Eq. (47) is greater or equal to

$$- \left(e^{2/3} \int \rho_{f_N}(x)^{4/3} dx \right)^{-1} E_{f_N} - \varepsilon.$$

Recalling Eq. (44), we have the result that $C_{N+1} \geq C_N - 2\varepsilon$. Since ε was arbitrary, $C_{N+1} \geq C_N$.

In the case of distinct e_i 's, one may define (for some fixed α , $0 < \alpha \leq 1$)

$$\begin{aligned} C_N(e_1, \dots, e_N) &= \sup_{f_N(x_1, \dots, x_N)} - \left[\left(\sum_{i=1}^N e_i^{1/2\alpha} \rho_{f_N}^i(x) \right)^{4/3} dx \right]^{-\alpha} \\ &\quad \times \left(\int \rho_{f_N}(x)^{4/3} dx \right)^{\alpha-1} E_{f_N}. \end{aligned}$$

A similar argument shows that these constants also increase, i.e., $C_N(e_1, \dots, e_N) \leq C_{N+1}(e_1, \dots, e_N, e_{N+1})$, where $e_{N+1} \geq 0$ is arbitrary. Of course Section 2 shows that $C_N(e_1, \dots, e_N) \leq 1.68$ for all N and e_i when $\alpha = \frac{1}{2}$.

Note added in proof:

In the text we proved the inequalities, Eqs. (6) and (7), when ψ is a wave function (pure state), and remarked that the inequalities also hold for a density matrix. To prove this, note that any density matrix, μ , can be written as $\mu = \sum_{\beta} \psi^{\beta} > < \psi^{\beta}$. In the definition, Eq. (4), simply regard β as just one more quantum number to sum over—on the same footing as the α 's. The rest of the proof is then the same as in the pure state case.

Bibliography

- [1] E. H. Lieb, *Phys. Lett.* **70A**, 444 (1979).
- [2] E. H. Lieb, *Rev. Mod. Phys.* **48**, 553 (1976).
- [3] E. H. Lieb and W. E. Thirring, *Phys. Rev. Lett.* **35**, 687 (1975); **35**, 1116 (1975) (errata).
- [4] P. A. M. Dirac, *Proc. Cambridge Philos. Soc.* **26**, 376 (1930).

- [5] S. R. Gadre, L. J. Bartolotti, and N. C. Handy, *J. Chem. Phys.* **72**, 1034 (1980).
- [6] F. Riesz, *J. London Math. Soc.* **5**, 162 (1930).
- [7] E. H. Lieb, *Stud. Appl. Math.* **57**, 93 (1977).
- [8] W. Feller, *An Introduction to Probability Theory and its Applications*, (Wiley, New York, 1966), Vol. 2, p. 261.
- [9] *British Association for the Advancement of Science Mathematical Tables*, (Office of the British Assoc., Burlington House, London, 1932) Vol. 2.

Received June 10, 1980.

Accepted for publication September 18, 1980

LETTER TO THE EDITOR

Monotonicity of the molecular electronic energy in the nuclear coordinates

Elliott H Lieb†

Departments of Mathematics and Physics, Princeton University, POB 708, Princeton, NJ 08544, USA

Received 3 December 1981

Abstract. Let $e(R_1, \dots, R_k)$ be the electronic contribution to the ground-state energy of a molecule consisting of one electron and k nuclei located at $R_1, \dots, R_k$. It is shown that $e(R'_1, \dots, R'_k) \geq e(R_1, \dots, R_k)$ if $|R'_i - R'_j| \geq |R_i - R_j|$ for all pairs i, j .

Consider the non-relativistic Schrödinger ground-state energy, E , of a molecule consisting of N electrons and k fixed nuclei with positive charges $z_1, \dots, z_k$ located at distinct points $R_1, \dots, R_k \in \mathbb{R}^3$. (The approximation that the nuclei are fixed is sometimes called the Born–Oppenheimer approximation.) Since we shall be interested in the dependence of E on the R_i , with the z_i fixed, we use the notation $\underline{R} = (R_1, \dots, R_k)$ to indicate the relevant variables. Further, we assume $\hbar^2/2m = 1$ and $e = -1$ where m and e are the electron mass and charge.

$$E(\underline{R}) = e(\underline{R}) + U(\underline{R}) \quad (1)$$

$$U(\underline{R}) = \sum_{1 \leq i < j \leq k} z_i z_j |R_i - R_j|^{-1}. \quad (2)$$

$U(\underline{R})$ is the nuclear–nuclear potential energy while $e(\underline{R})$ is the electronic contribution to the energy and is the ground-state energy (inf spectrum) of the Hamiltonian

$$H(\underline{R}) = - \sum_{i=1}^N (\Delta_i + V(x_i, \underline{R})) + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} \quad (3)$$

with

$$V(x, \underline{R}) = \sum_{j=1}^k z_j |x - R_j|^{-1} \quad (4)$$

being the negative of the electron–nuclear interaction.

An old conjecture is that when the R_i are moved apart, in some sense that has to be made precise, and if N is not too large, $e(\underline{R})$ *increases*. ($U(\underline{R})$ obviously *decreases*.) In other words $e(\underline{R})$ is attractive. Thus, binding (i.e. a minimum for $E(\underline{R})$) could be viewed as a competition between the electronic attraction and the nuclear–nuclear repulsion.

† Work partially supported by US National Science Foundation Grant PHY-7825390 A02.

L64 *Letter to the Editor*

Three years ago, Lieb and Simon (1978) proved[†] that $e(\underline{R})$ is indeed attractive when $N = 1$ (and for any choice of the positive charges z_i) in the following sense:

Theorem 1. Let $\underline{R}$ and $\underline{R}'$ be two sets of nuclear coordinates such that for some $\lambda > 1$ and all i , $\underline{R}'_i = \lambda \underline{R}_i$ (i.e. uniform dilation). Then $e(\underline{R}) \leq e(\underline{R}')$.

The reader is referred to Lieb and Simon (1978) for a discussion of the problems that arise in trying to extend this theorem to $N > 1$, and for a proof of the fact that for all N , $e(\underline{R})$ achieves its minimum value when $R_1 = R_2 = \dots = R_k$.

The main point of the present letter is theorem 2. For $N = 1$, $e(\underline{R})$ is attractive for a more general class of deformations of $\underline{R}$ than simple dilation. This larger class is quite natural, for it is more nearly the class of deformations for which $U(\underline{R})$ is repulsive (i.e. $U(\underline{R})$ decreases).

Definition 1. Let $\underline{R}$ and $\underline{R}'$ be two sets of k coordinates in $\mathbb{R}^3$. $\underline{R}'$ is said to be *bigger than* $\underline{R}$, ($\underline{R}' > \underline{R}$) if $|\underline{R}'_i - \underline{R}'_j| \geq |\underline{R}_i - \underline{R}_j|$ for every pair (i, j) of coordinates. (Note: It is *not* necessary that there be a continuous deformation of $\underline{R}$ into $\underline{R}'$ such that the pairwise distances $|\underline{R}_i - \underline{R}_j|$ increase during the deformation. The stated inequality is the only condition.)

Clearly,

$$\underline{R}' > \underline{R} \Rightarrow U(\underline{R}') \leq U(\underline{R}). \quad (5)$$

Theorem 2. Let $N = 1$ and let the $z_i > 0$ be arbitrary but fixed. If $\underline{R}'$ is bigger than $\underline{R}$ then $e(\underline{R}') \geq e(\underline{R})$.

Remark. Obviously the electron–electron repulsion $|x_i - x_j|^{-1}$ plays no role in these theorems since $N = 1$. The only relevant function is the elementary electron–nuclear potential, $w(x)$, which in our case is $w(x) = |x|^{-1}$. Theorem 1 holds if $|x|^{-1}$ in equation (4) is replaced by any w satisfying the following condition:

(C1). For every real number s , the set $B_s = \{x | w(x) > s\}$ is convex and balanced (i.e. $x \in B_s \Rightarrow -x \in B_s$). In particular, theorem 1 holds under the following condition.

(C2). The function $w(x)$ is spherically symmetric and is a decreasing function of $r = |x|$.

While it is plausible that theorem 2 holds under condition (C2) we are unable to prove it by the methods given here. The sort of problems that arises in following the ‘log concavity’ proof of theorem 1 in Lieb and Simon (1978) is similar to the problem of proving the Klee (1979) conjecture about the intersection of spheres.

The proof of theorem 2 given here does show that theorem 2 holds if $|x|^{-1}$ is replaced by any w satisfying both of the following more stringent conditions:

(C3). The function w is spherically symmetric.

(C4). If $u(r) \equiv w(r^{1/2})$ then $u(r)$ is completely monotone (see definition 2). For the Coulomb case, $u(r) = r^{-1/2}$ and this is completely monotone. More generally, each of the k potentials in equation (4) can be replaced by a different w satisfying (C3) and (C4).

[†] A different proof for the case of two nuclei was later given in Hoffmann-Ostenhof T 1980 *J. Phys. A: Math. Gen.* **13** 417–24.

Definition 2. A function $u(r)$, defined for $0 < r < \infty$, is *completely monotone* if u is infinitely differentiable and $(-1)^j u^{(j)}(r) \geq 0$ for all r and all $j = 0, 1, 2, \dots$ (with $u^{(j)}$ denoting the j th derivative).

The following lemma due to Bernstein (1928) (see also Donoghue 1974) is required for the proof of theorem 2.

Lemma 3. The function $u(r)$ is completely monotone if and only if $u(r)$ is the (one-sided) Laplace transform of a positive measure, i.e.

$$u(r) = \int \exp(-tr) d\mu(t) \quad (6)$$

where μ is a (positive) measure supported on $[0, \infty)$.

Clearly, if $u(r)$ is completely monotone then so is $\exp(bu(r))$ for every $b \geq 0$. Lemma 3 thus yields the following representation for all $b \geq 0$

$$\exp(b|x - R|^{-1}) = \int \exp(-t|x - R|^2) d\mu(t; b). \quad (7)$$

A similar representation holds if $|x - R|^{-1}$ is replaced by $w(x - R)$ on the left side of equation (7). The reader may wish to have an explicit formula for μ in equation (7). This is provided by the formulae (12.2.1) on p 72 and (29) on p 171 in Roberts and Kaufman (1966)

$$d\mu(t; b) = \left(t^{-3/2} (b/\pi)^{1/2} \int_0^\infty u^2 I_1(2b^{1/2}u) \exp(-u^4/4t) du + \delta(t) \right) dt$$

where $\delta(t) dt$ is the Dirac delta measure.

The second tool that will be needed is the following simple lemma.

Lemma 4. Let $A = \{A_{ij}\}$ be a positive-semidefinite real symmetric N -square matrix satisfying (i) $\sum_{i=1}^N A_{ij} = 0$ and (ii) $A_{ij} \leq 0$ when $i \neq j$. Let $M > 0$ be an integer and let R_{ij} ($i = 1, \dots, N$ and $j = 1, \dots, M$) be NM vectors in $\mathbb{R}^3$ (the choice of three dimensions is made for simplicity). Denote these collectively by $\underline{R}$. Let $\lambda_{ij} \geq 0$ be a collection of NM real numbers satisfying $\Lambda_i \equiv \sum_{j=1}^M \lambda_{ij} > 0$. Let $x_1, \dots, x_N \in \mathbb{R}^3$ be N three-dimensional variables. Let

$$Z(\underline{R}) \equiv \int dx_1 \dots dx_N \exp \left(- \sum_{i,j=1}^N x_i \cdot x_j A_{ij} - \sum_{n=1}^N \sum_{j=1}^M \lambda_{ij} |x_i - R_{ij}|^2 \right). \quad (8)$$

If $\underline{R}' > \underline{R}$ (i.e. $|R'_{ij} - R'_{kl}| \geq |R_{ij} - R_{kl}|$ for all i, j, k, l) then $Z(\underline{R}') \leq Z(\underline{R})$.

Proof. The integral for Z is Gaussian and easily evaluated. Let Λ denote the diagonal matrix with the values Λ_i . By (i), $\sum_{j=1}^N (A + \Lambda)_{ij}^{-1} \Lambda_j = 1$ (since $\sum_j (A + \Lambda)_{ij} = \Lambda_i$). Using this fact, a little algebra shows that

$$Z(\underline{R}) = \pi^{3N/2} \exp \left(-\frac{1}{2} \sum_{i,j=1}^N \sum_{k,l=1}^M (A + \Lambda)_{ij}^{-1} \lambda_{ik} \lambda_{jl} |R_{ik} - R_{jl}|^2 - \frac{3}{2} \ln \text{Det}(A + \Lambda) \right).$$

The lemma will be proved if $(A + \Lambda)_{ij}^{-1} \geq 0$ for all i, j . But (ii) states that $A + \Lambda$ has negative off-diagonal matrix elements. The inverse of a positive definite matrix P with this property has $P_{ij}^{-1} \geq 0$ (Ostrowski 1937). QED.

With this preparation we can now prove theorem 2. In fact, something stronger will be proved.

Theorem 5. For $T > 0$ let $G(x, y, T, \underline{R}) = \exp(-TH(\underline{R}))(x, y)$ be the kernel of the operator $\exp(-TH(\underline{R}))$. Let $(\underline{R}, x, y)$ denote the $k+2$ vectors $R_1, \dots, R_k, x, y$. Suppose that $(\underline{R}', x', y')$ is bigger than $(\underline{R}, x, y)$. Then $G(x' y', T, \underline{R}') \leq G(x, y, T, \underline{R})$.

Proof. First, regularise $w(x) = |x|^{-1}$ by $w_\epsilon(x) = w((x^2 + \epsilon)^{1/2}) = (x^2 + \epsilon)^{-1/2}$ for $\epsilon > 0$. This amounts to replacing $d\mu(t; b)$ by $\exp(-\epsilon t) d\mu(t; b)$ in equation (7). Second, use the Trotter product formula (see Lieb and Simon 1978)

$$G(x, y, T, \underline{R}) = \lim_{n \rightarrow \infty} [\exp(T\Delta/n) \exp(TV(\underline{R})/n)]^n(x, y)$$

where $V(\underline{R})$ is given in equation (4) (with w_ϵ).

$$\exp(T\Delta/n)(x, y) = (4\pi T/n)^{-3/2} \exp[-n(x-y)^2/4T]$$

is a Gaussian with negative off-diagonal matrix elements (i.e. $n/4T$). The total Gaussian has the exponent $\alpha + \beta$ with

$$\alpha = -(n/4T) \sum_{i=1}^{n-2} (x_{i+1} - x_i)^2$$

$$\beta = -(n/4T)[(x_1 - x)^2 + (x_{n-1} - y)^2]$$

where the x_i are the intermediate integration variables. The first term, α , should be regarded as the 'A' term in equation (8). The second term, β , should be regarded as part of the second term in equation (8). Use the representation, equation (7), nk times. The theorem (for w_ϵ) then follows from lemma 4. Finally, we let $\epsilon \rightarrow 0$. (Remark: $G(x, y, T, \underline{R})$ is monotone decreasing in ϵ since w_ϵ is decreasing.) QED.

Corollary 6. For the simple hydrogen atom (with $V(x) = |x|^{-1}$), $G(x', y', T) \leq G(x, y, T)$ if $|x'| \geq |x|$, $|y'| \geq |y|$ and $|x' - y'| \geq |x - y|$.

Proof of theorem 2. Take $y = x$, $y' = x'$ and choose these such that $|x' - R'_i| \geq |x - R_i|$, $i = 1, \dots, k$. Since $e(\underline{R}) = -\lim_{T \rightarrow \infty} T^{-1} \ln G(x, x, T, \underline{R})$ (for any x), theorem 5 implies theorem 2. QED.

References

- Bernstein S 1928 *Acta Math.* **52** 1–66
 Donoghue W F Jr 1974 *Monotone Matrix Functions and Analytic Continuation* (New York: Springer) p 13
 Klee V 1979 *Math. Mag.* **52** 131–145 (see problem 7 on p 142)
 Lieb E H and Simon B 1978 *J. Phys. B: At. Mol. Phys.* **11** L537–42
 Ostrowski A 1937 *Comm. Math. Helv.* **10** 69–96
 Roberts G E and Kaufman H 1966 *Table of Laplace Transforms* (Philadelphia: Saunders)

Proof of the Stability of Highly Negative Ions in the Absence of the Pauli Principle

Rafael Benguria

Departamento de Física, Universidad de Chile, Casilla 5487, Santiago, Chile

and

Elliott H. Lieb

Departments of Mathematics and Physics, Princeton University, New Jersey 08544

(Received 31 January 1983)

It is well known that ionized atoms cannot be both very negative and stable. The maximum negative ionization is only one or two electrons, even for the largest atoms. The reason for this phenomenon is examined critically and it is shown that electrostatic considerations and the uncertainty principle cannot account for it. The exclusion principle plays a crucial role. This is shown by proving that when Fermi statistics is ignored, then the degree of negative ionization is at least of order z , the nuclear charge, when z is large.

PACS numbers: 31.10.+z, 03.65.Ge

One of the interesting and important facts about atoms is that they cannot be very negatively ionized (in a stable state, as distinguished from metastable state). For a nucleus of charge z , let $N_c(z)$ denote the maximum number of electrons that can be bound to this nucleus (*in vacuo*, not in water or other matter). Experiments indicate that $N_c(z) - z$ is one, or possibly two, as z varies over the periodic table. It is often said that this striking fact, which begs for an explanation, is a consequence of electrostatics; namely, if an atom has a net negative charge then an additional electron will not bind because the electron can lower its energy by escaping to infinity. The purpose of this note is to examine this simple, but important physical problem in a critical way and to show that the correct explanation does not lie with electrostatics alone—the *Pauli exclusion principle plays a central role in the correct explanation*. We are not able to offer an explanation of the phenomenon, but we thought it worthwhile at least to expose the fallacy in the “simple electrostatic” explanation and thereby show that the phenomenon is really a deep consequence of quantum mechanics.

To prove that the Pauli principle is essential we shall consider an atom in which the electrons are spinless bosons. (Since the ground state of a many-body system is nodeless, it is automatically symmetric; therefore “bosons” and “ignoring statistics” are synonymous.) We shall prove that in this model, $N_c(z) - z \geq \gamma z$ when z is large, and where $\gamma > 0$ is some fixed constant. We do not know the numerical value of γ , except that $0 < \gamma < 1$. It can be found by solving an equation [namely (11) with $\mu=0$] on a computer, if there is suffi-

cient interest in doing so. The exact numerical value of γ is not as important as the fact that “bosonic” atoms would not obey the $N_c(z) - z \approx 1$ rule for sufficiently large z . Just how large z has to be in order to violate the rule substantially, we do not know. Equation (11) has to be solved to answer the question.

One can adopt different points of view about this. It is possible that the rule is not really a rule at all for fermions, and that $N_c(z) - z$ grows at least as fast as z for large z . In this case the fact that $N_c(z) - z \approx 1$ within the periodic table is fortuitous, and in reality bosons and fermions are qualitatively similar as far as the phenomenon goes. Another possibility is that $N_c(z)/z - 1 \rightarrow 0$ as $z \rightarrow \infty$ [thereby allowing the possibility that $N_c(z) - z \approx z^{1/2}$, for example], in which case the Pauli principle is crucial. We, of course, do not know which point of view will ultimately prevail. It is to be hoped that if it is the second one then someone will find a simple, but rigorous explanation. In any case, the phenomenon should not be left merely as a numerical statement about the periodic table but should be understood on a deeper level.

While we use the nonrelativistic Schroedinger equation and we regard the nucleus as fixed, it will be clear, at least intuitively, that our conclusions are not limited by these approximations. The lack of the Pauli principle is, however, crucial.

Before turning to the mathematical proof, let us consider the problem from a heuristic viewpoint. Suppose N electrons are bound to the nucleus. With neglect of many-body effects, the effective potential that an $(N+1)$ th electron feels

is approximately

$$\varphi_\rho(\vec{x}) = V(\vec{x}) + \int |\vec{x} - \vec{y}|^{-1} \rho(\vec{y}) d^3y, \quad (1)$$

where $V(\vec{x}) = -z/|\vec{x}|$, and units in which the electron charge is unity are used. $\rho(\vec{x})$ is the density of the N electrons, $\int \rho = N$. The effective Hamiltonian for the $(N+1)$ th electron is approximately

$$h_{\text{eff}} = -m^{-1}\Delta + \varphi_\rho(\vec{x}), \quad (2)$$

where m is the electron mass and $\hbar^2 = 2$.

As N increases from zero, φ_ρ increases in some average sense. When $N > z$, $\varphi_\rho(\vec{x})$ is positive for large $|\vec{x}|$ [by Newton's theorem, $\varphi_\rho(\vec{x}) \approx (N-z)/|\vec{x}|$ for large $|\vec{x}|$]. However, $\varphi_\rho(\vec{x})$ is still very large and negative for x near zero, namely, $-z/|\vec{x}|$. The uncertainty principle is crucial here; it prevents $\rho(\vec{x})$ from being a delta function and thereby screening the nucleus for small $|\vec{x}|$. Thus, even if $N > z$, h_{eff} might have a genuine bound state and the $(N+1)$ th electron might be bound. Eventually, of course, the region of negative φ_ρ will be too small and binding will cease. Implicit in this discussion is the fact that the $(N+1)$ th electron is allowed to go into any available bound state of h_{eff} . In other words, the argument works if statistics is ignored, which is the same thing as saying that we are dealing with bosons. If, on the other hand, the electrons are fermions then, for binding, h_{eff} must have something like $N+1$ bound states in order that the $(N+1)$ th electron can go into an orbital that is orthogonal to the previously occupied orbitals. It is a remarkable fact about Fermi statistics that the $(N+1)$ th bound state of h_{eff} disappears when $N \approx z$, if the $N_c(z) - z \approx 1$ rule is obeyed.

The above argument is not completely convincing, even on the heuristic level, because it is not obvious that h_{eff} indeed has a bound state when $N = (1+\gamma)z$.

A proper proof, starting from the correct Schrodinger equation, and without any approximations, will now be given. Let

$$H(N) = \sum_{i=1}^N [-m^{-1}\Delta_i + V(\vec{x}_i)] + \sum_{1 \leq i < j \leq N} |\vec{x}_i - \vec{x}_j|^{-1} \quad (3)$$

be the Hamiltonian for N bosonic electrons [$V(\vec{x}) = -z/|\vec{x}|$]. Let $E(N)$ be the ground-state energy

of $H(N)$. We shall find two bounding functions, $E_+(N)$, with $E_-(N) \leq E(N) \leq E_+(N)$ and

$$E_+(N) = -mz^3 e(N/z), \quad (4a)$$

$$E_-(N) = -mz^3 [e(N/z)^{1/2} + bz^{-3/2} N^{5/6}]^2. \quad (4b)$$

Here, $b = 0.36$ and $e(t)$ is a monotone nondecreasing, concave function ($\dot{e} \geq 0$, $\ddot{e} \leq 0$), defined for all real $t \geq 0$, with $e(0) = 0$. $E_+(N)$ is essentially the Hartree energy. The crucial point about $e(t)$ is that it is strictly increasing in t up to some $t_c = 1 + \gamma$ and $0 < \gamma < 1$. We do not know the numerical value of γ , but that is unimportant. [It can be found by solving Eq. (11), as explained below.] From Eq. (4) we see that if $N < \bar{N}$, where

$$E_-(\bar{N}) = E_+(zt_c) \quad (5)$$

(which means that $\bar{N}$ is not necessarily an integer), then the energy can be lowered by adding some number of electrons because $E(N) \geq E_-(N) > E_-(\bar{N}) = E_+(zt_c) = E_+([zt_c]) \geq E([zt_c])$. Here, $[x]$ denotes the smallest integer $\geq x$. [Note that Eq. (5) has a unique solution; $E_-(N)$ is monotone in N since $e(N/z)$ is monotone in N .] In other words

$$N_c(z) \geq [\bar{N}]. \quad (6)$$

As $z \rightarrow \infty$, $\bar{N} \rightarrow z(1+\gamma)$ because (5) reads (with $\bar{N}/z \equiv t$)

$$e(t)^{1/2} + bz^{-2/3} t^{5/6} = e(t_c)^{1/2}. \quad (7)$$

Since the solution to Eq. (7) satisfies $t < t_c < 2$ and since $e(t)$ is continuous, we have that $t \rightarrow t_c$ as $z \rightarrow \infty$. Thus, our main point is proved, namely, asymptotically

$$N_c(z) - z \geq \gamma z, \text{ large } z. \quad (8)$$

An upper bound for $N_c(z)$ was first given by Ruskai in the form $N_c(z) \leq (\text{const})z^2$ for bosons.¹ Sigal² proved for fermions that $N_c(z) \leq cz$ for z sufficiently large, with $c > 2$ being some constant. For fermions Ruskai³ proved that $N_c(z) \leq (\text{const}) \times z^{6/5}$. Sigal⁴ improved his result for fermions to $N_c(z) \leq \alpha(z)z$, with $\alpha(z) \leq 12$ and $\alpha(z) \rightarrow 2$ as $z \rightarrow \infty$.

Now we turn to the proof of Eq. (4).

Upper bound for $E(N)$.—We take a product trial function $\psi = \prod_{i=1}^N f(\vec{x}_i)$, with $\int f^2 = 1$ and with $f(\vec{x})$ real and nonnegative, and use the variational principle:

$$E(N) \leq \langle \psi | H(N) | \psi \rangle \leq \int \{ m^{-1} |\nabla \rho(\vec{x})|^2 + V(\vec{x}) \rho(\vec{x}) \} d^3x + \frac{1}{2} \iint \rho(\vec{x}) \rho(\vec{y}) |\vec{x} - \vec{y}|^{-1} d^3x d^3y \equiv L(\rho), \quad (9)$$

where $\rho(\tilde{\mathbf{x}}) \equiv Nf(\tilde{\mathbf{x}})^2$. [We could insert a factor $(N-1)/N$ before the last integral, but choose not to do so because we want $L(\rho)$ to be independent of N .] Next we define

$$E_+(N) = \inf \{ L(\rho) \mid \int \rho = N, \rho(\tilde{\mathbf{x}}) \geq 0 \}. \quad (10)$$

Equation (10) means that we try to minimize $L(\rho)$ under the stated conditions. The minimum may not be achieved by any ρ (it is achieved if and only if $N \leq 1 + \gamma$ as stated below), but in any case E_+ cannot exceed the "greatest lower bound" or "infimum" of $L(\rho)$. The problem posed by Eq. (10) is a special case of the generalized Thomas-Fermi-von Weizsaecker problem analyzed earlier.^{5,6} In our case, by the simple scaling $\rho(\tilde{\mathbf{x}}) \rightarrow m^3 z^4 \rho(mz\tilde{\mathbf{x}})$ the problem can be reduced to the case in which $m=1$, $z=1$, and $\int \rho = N/z$. Equation (4a) is thus seen to hold. Moreover, the m dependence of e is $e(N/z, m) = me(N/z, m=1)$ as in Eq. (4a). (Incidentally, this scaling shows that the radius of bosonic atoms *shrinks* as z^{-1} , whereas it *shrinks* as $z^{-1/3}$ for fermions.) There is a minimizing ρ for $L(\rho)$ (and it is unique) if and only if $N/z \leq t_c$ for some definite number $t_c > 1$. (Note that t_c is independent of m .) This ρ

satisfies

$$[-m^{-1}\Delta + \varphi_\rho(\tilde{\mathbf{x}})]\rho(\tilde{\mathbf{x}})^{1/2} = -\mu\rho(\tilde{\mathbf{x}})^{1/2} \quad (11)$$

with $\mu \geq 0$, and $\mu = 0$ when $N = t_c z$. The proof that $t_c > 1$ is given in Ref. 5, lemma 13, and Ref. 6, theorem 7.16 (note: in these proofs take $p = \frac{5}{3}$ and $\gamma = 0$). The proof that $t_c < 2$ is given in Ref. 6, theorem 7.23. [The reason that $t_c > 1$ is that when $N = z$ then $\varphi_\rho(\tilde{\mathbf{x}}) < 0$ and one can prove that this potential has a bound state; thus μ cannot be zero. To prove that $t_c < 2$, set $\mu = 0$, multiply Eq. (11) by $|\tilde{\mathbf{x}}|\rho(\tilde{\mathbf{x}})^{1/2}$ and integrate. One can show, by partial integration, that $\int |\tilde{\mathbf{x}}|\rho(\tilde{\mathbf{x}})^{1/2}\Delta\rho(\tilde{\mathbf{x}})^{1/2} \leq 0$. Obviously, $\int V(\tilde{\mathbf{x}})|\tilde{\mathbf{x}}|\rho(\tilde{\mathbf{x}}) = -zN$. Finally, $I \equiv \int \int |\tilde{\mathbf{x}}|\rho(\tilde{\mathbf{x}})\rho(\tilde{\mathbf{y}})|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}|^{-1} = \frac{1}{2} \int \int \rho(\tilde{\mathbf{x}})\rho(\tilde{\mathbf{y}})|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}|^{-1}(|\tilde{\mathbf{x}}| + |\tilde{\mathbf{y}}|)$. But $(|\tilde{\mathbf{x}}| + |\tilde{\mathbf{y}}|)|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}|^{-1} \geq 1$, so that $I \geq N^2/2$. Since $I \leq zN$, $N \leq 2z$.] That $e(t)$ is concave (and strictly concave when $t \leq t_c$) is a consequence of the fact that when $\int \rho$ is increased the additional density can be placed, if need be, at infinity where its energy contribution is zero (see Refs. 5 and 6).

Lower bound for $E(N)$.—Let $\psi(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N)$ be any normalized function. We want to show that

$$F(\psi) \equiv \langle \psi | H(N) | \psi \rangle \geq \text{right side of (4b)}. \quad (12)$$

Let

$$\rho_\psi(\tilde{\mathbf{x}}) = \sum_{i=1}^N \int |\psi(\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{i-1}, \tilde{\mathbf{x}}, \tilde{\mathbf{x}}_{i+1}, \dots, \tilde{\mathbf{x}}_N)|^2 d^3x_1 \dots d^3x_{i-1} d^3x_{i+1} \dots d^3x_N \quad (13)$$

be the single-particle density associated with ψ ; $\int \rho_\psi = N$. We shall use several known inequalities. The first is the kinetic energy inequality of Hoffmann-Ostenhof⁷ (see also Ref. 8 for a further discussion of kinetic energy inequalities):

$$\langle \psi | -\sum_{i=1}^N \Delta_i | \psi \rangle \geq \int |\nabla \rho_\psi(\tilde{\mathbf{x}})|^{1/2}|^2 d^3x. \quad (14)$$

This follows by taking the gradient in (13) and then using the Schwarz inequality. The second is the "exchange and correlation" inequality^{9,10}:

$$\langle \psi | \sum_{1 \leq i < j \leq N} |\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j|^{-1} | \psi \rangle \geq \frac{1}{2} \int \int \rho_\psi(\tilde{\mathbf{x}})\rho_\psi(\tilde{\mathbf{y}})|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}|^{-1} d^3x d^3y - (1.68) \int \rho_\psi(\tilde{\mathbf{x}})^{4/3} d^3x. \quad (15)$$

Inserting Eqs. (14) and (15) in (12) we have, for any ψ ,

$$F(\psi) \geq L(\rho_\psi) - (1.68) \int \rho_\psi(\tilde{\mathbf{x}})^{4/3} d^3x. \quad (16)$$

To bound the right side of (16) from below, let us first explicitly denote the m dependence of $L(\rho)$ by $L(\rho, m)$. Choose $\epsilon > 0$ and let $m_\epsilon = (1 + \epsilon)m$. Then Eq. (16) reads

$$F(\psi) \geq L(\rho_\psi, m_\epsilon) + P(\rho_\psi, m_\epsilon), \quad (17)$$

$$P(\rho_\psi, m_\epsilon) = \epsilon m_\epsilon^{-1} \int |\nabla \rho_\psi(\tilde{\mathbf{x}})|^{1/2}|^2 d^3x - (1.68) \int \rho_\psi(\tilde{\mathbf{x}})^{4/3} d^3x. \quad (18)$$

We have already seen that [see Eq. (10) and the following remark about scaling]

$$L(\rho_\psi, m_\epsilon) \geq (m_\epsilon/m)E_+(N) = (1 + \epsilon)E_+(N). \quad (19)$$

To bound P , we use the Sobolev inequality^{6,11}

$$\int |\nabla g(\tilde{x})|^2 d^3x \geq 3(\pi/2)^{4/3} \left\{ \int |g(\tilde{x})|^6 d^3x \right\}^{1/3} \quad (20)$$

for any g . Thus

$$P(\rho_\psi, m_\epsilon) \geq 3(\pi/2)^{4/3} \epsilon m_\epsilon^{-1} \left\{ \int \rho_\psi(\tilde{x})^3 d^3x \right\}^{1/3} - (1.68) \int \rho_\psi(\tilde{x})^{4/3} d^3x. \quad (21)$$

By Hoelder's inequality, $\int \rho^{4/3} \leq X \left\{ \int \rho^{5/6} \right\}^{8/5}$ with $X = \left\{ \int \rho^3 \right\}^{1/6}$. Inserting this in (21), and then minimizing the right side with respect to the unknown X , we have

$$P(\rho_\psi, m_\epsilon) \geq -b^2 m_\epsilon N^{5/3} / \epsilon, \quad (22)$$

$$b = (1.68) 2^{-1/3} 3^{-1/2} \pi^{-2/3} = 0.36. \quad (23)$$

Inserting Eqs. (19) and (22) in (17) we have, for any $\epsilon > 0$, and any normalized ψ ,

$$\langle \psi | H(N) | \psi \rangle \geq (1 + \epsilon) E_+(N) - (1 + 1/\epsilon) m b^2 N^{5/3}. \quad (24)$$

Maximizing this with respect to ϵ yields Eq. (4b).

One of the authors (E.L.) acknowledges gratefully the hospitality and support of the Departamento de Física, Universidad de Chile, where this work was carried out. He also acknowledges the support of the U. S. National Science Foundation under Grant No. PHY-8116101 A01. We thank Barry Simon for valuable comments.

¹M. B. Ruskai, Commun. Math. Phys. **82**, 457 (1982).

²I. M. Sigal, Commun. Math. Phys. **85**, 309 (1982). See also *Mathematical Problems in Theoretical Physics*, Lecture Notes in Theoretical Physics Vol. 153 (Springer-Verlag, Berlin, 1982), pp. 149-156.

³M. B. Ruskai, Commun. Math. Phys. **85**, 325 (1982).

⁴I. M. Sigal, Institute Mittag-Leffler Report No. 12, 1982 (unpublished).

⁵R. Benguria, H. Brezis, and E. H. Lieb, Commun. Math. Phys. **79**, 167 (1981).

⁶E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981).

⁷M. Hoffmann-Ostenhof and T. Hoffmann-Ostenhof, Phys. Rev. A **16**, 1782 (1977).

⁸E. H. Lieb, *Density Functionals for Coulomb Systems*, in *Physics as Natural Philosophy: Essays in Honor of Laszlo Tisza on His 75th Birthday*, edited by A. Shimony and H. Feshbach (MIT Press, Cambridge, Mass., 1982), pp. 111-149.

⁹E. H. Lieb, Phys. Lett. **70A**, 444 (1979).

¹⁰S. Oxford and E. H. Lieb, Int. J. Quantum Chem. **19**, 427 (1981).

¹¹E. H. Lieb, Rev. Mod. Phys. **48**, 553 (1976).

PHYSICAL REVIEW LETTERS

VOLUME 52

30 JANUARY 1984

NUMBER 5

Atomic and Molecular Negative Ions

Elliott H. Lieb

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08544
(Received 16 November 1983)

An upper bound is given for the maximum number, N_c , of negative particles (fermions or bosons or a mixture of both) of charge $-e$ that can be bound to an atomic nucleus of charge $+ze$. If z is integral then $N_c \leq 2z$. In particular, this is the first proof that H^{--} is not stable. For a molecule, $N_c \leq 2Z + K - 1$, where K is the number of atoms in the molecule and Z is the total nuclear charge.

PACS numbers: 03.65.Ge, 31.10.+z

One of the striking, nonperiodic facts about the periodic table is that the maximum number of electrons, N_c , that can be bound to a nucleus of charge z is never more than $z + 1$. Recently, several authors¹⁻⁸ have attempted to find bounds on N_c ; one of the strongest results so far⁸ (for fermions) is that $\lim_{z \rightarrow \infty} N_c/z = 1$. For *bosons*,⁵ however, $N_c > 1.2z$ for large z . Thus, the value of N_c is very dependent on the statistics of the bound particles.

The purpose of this note is to announce a theorem about N_c , the full details of which will appear elsewhere.⁹ The theorem applies to *any mixture* of bound particles, with possibly different statistics, masses, and charges (as long as they are all negative), and even with possibly different magnetic fields acting on the various particles. (Naturally, symmetry requires that particles of the same type have the same mass, etc.) The theorem also applies to a molecule. The usual approximation that the nuclei be fixed (or infinitely massive) is important, but if they are not fixed a weaker theorem holds. The same theorems hold in the Hartree-Fock (restricted or unrestricted) and Hartree approximations to the ground-state energy.⁹

Suppose that we have a molecule with K nuclei of charges $z_1, \dots, z_K > 0$ (units are used in which the electron charge is unity) located at fixed, dis-

tinct positions $\vec{R}_1, \dots, \vec{R}_K$. The electric potential of these nuclei is

$$V(\vec{x}) = \sum_{j=1}^K z_j |\vec{x} - \vec{R}_j|^{-1}. \quad (1)$$

Let there be N negative particles with masses $m_1, \dots, m_N$ and charges $-q_1, \dots, -q_N < 0$ (in the usual case each $q_i = 1$) and let each be subject to (possibly different) magnetic fields $\vec{A}_1(\vec{x}), \dots, \vec{A}_N(\vec{x})$. (The generality of allowing nonintegral nuclear and negative particle charges may have some physical relevance because, as pointed out to me by W. Thirring, particles in solids such as semiconductors may have nonintegral effective charges due to dielectric effects.) The Hamiltonian is

$$H_N = \sum_{j=1}^N \{T_j - q_j V(\vec{x}_j)\} + \sum_{1 \leq i < j \leq N} q_i q_j |\vec{x}_i - \vec{x}_j|^{-1}. \quad (2)$$

Here, T_j is the kinetic energy operator for the j th particle and it is one of the following (possibly different for different j) two types (nonrelativistic or relativistic):

$$T_j = [\vec{p}_j - q_j \vec{A}_j(\vec{x})/c]^2 / 2m_j, \quad (3)$$

$$T_j = \{[\vec{p}_j c - q_j \vec{A}_j(\vec{x})]^2 + m_j^2 c^4\}^{1/2} - m_j c^2. \quad (4)$$

Let q denote the maximum of the q_j , let $Q = \sum_{j=1}^N q_j$ be the total negative charge, and let $Z = \sum_{j=1}^K q_j$ be the total nuclear charge. Let E_N denote the ground-state energy of $H_N \equiv \inf \text{spec}(H_N)$.

Theorem 1.—If the above system is bound (meaning that E_N is an eigenvalue of H_N) then, necessarily,

$$Q < 2Z + qK. \quad (5)$$

In the atomic case ($K=1$) this can be strengthened to

$$Q < 2Z + \sum_{j=1}^N q_j^2/Q. \quad (6)$$

The strict inequality in Eqs. (5) and (6) is important; in the atomic case with $q=1$ and z integral, Eq. (5) implies

$$N_c \leq 2z. \quad (7)$$

For a hydrogen atom ($z=1$), Eq. (7) implies that $N_c=2$ (since it is known that two electrons can, in fact, be bound). H^- is *not* stable. This result had not been proved before, although there exist partial results in this direction.⁶

Although Eq. (7) is far from optimal when z is large [in view of the $N \approx z + O(1)$ conjecture], it is the strongest explicit estimate obtained so far and it is of the right order of magnitude for bosons (recalling the $N_c > 1.2z$ result⁵).

The theorem holds even if the nuclei are not points, but are spherical charge distributions, i.e., $|\vec{x} - \vec{R}|^{-1}$ is replaced by $\int d\mu(\vec{y}) |\vec{y} - \vec{x} - \vec{R}|^{-1}$ in Eq. (1), where μ is any positive, spherical measure of unit total charge.

If the nuclear coordinates, $\vec{R}_j$, are dynamical instead of fixed, a weaker theorem holds. Let $\tilde{H}_N = H_N + T_{\text{nuc}} + U(\vec{R})$, where T_{nuc} is the nuclear kinetic energy [consisting of terms of the form in Eqs. (3) and (4)] and $U(\vec{R})$ is a potential depending on the nuclear coordinates. Let $\tilde{E}_N$ be the ground-state energy of $\tilde{H}_N$ and let $\tilde{E}_{N,j}^\infty$ be the ground-state energy when the nuclear masses are infinite (i.e., T_{nuc} is omitted) and the negative particle j is removed. If $\tilde{E}_{N,j}$ is the ground-state energy when particle j is removed, but T_{nuc} is retained, it is easy to see that $\tilde{E}_{N,j} \geq \tilde{E}_N$ and $\tilde{E}_{N,j} \geq \tilde{E}_{N,j}^\infty$. The theorem for dynamical nuclei, which assumes an additional inequality, is the following.

Theorem 2.—If the N -particle system is bound and if, in addition, $\tilde{E}_N \leq \tilde{E}_{N,j}^\infty$ for all $j=1, \dots, N$, then Eq. (5) [respectively Eq. (6) for $K=1$] holds.

The proof of the theorems is simple enough to be given in an elementary quantum mechanics

course—at least in the atomic case with fixed nucleus and with $T_j = p_j^2/2m_j$ (all j). The proof (ignoring some technical fine points) in this atomic case is the following: Take $\hbar^2/2m = 1$ and $\vec{R} = 0$. Pick some j and write $H_N = H_{N,j} + h_j$, where $H_{N,j}$ is the Hamiltonian for the remaining $N-1$ particles and

$$h_j = T_j - q_j z |\vec{x}_j|^{-1} + \sum_{k \neq j} |\vec{x}_j - \vec{x}_k|^{-1} q_j q_k. \quad (8)$$

Assume that the system is bound and let ψ be the ground state (which is real). Multiply the Schrödinger equation, $H_N \psi = E_N \psi$, by $|\vec{x}_j| \psi$ and integrate over all N variables. Let X_j denote all the $N-1$ variables other than $\vec{x}_j$. For the $H_{N,j}$ term, do the $d^{3(N-1)}X_j$ integration first; by the variational principle, the X_j integral is, for each fixed $\vec{x}_j$, not less than $E_{N,j}$ ($\equiv$ the ground-state energy of $H_{N,j}$) times the same integral without $H_{N,j}$. This inequality is preserved after the $\vec{x}_j$ integration since $|\vec{x}_j|$ is a positive weight. Thus

$$\langle |\vec{x}_j| \psi | H_{N,j} \psi \rangle \geq E_{N,j} \langle |\vec{x}_j| \psi | \psi \rangle. \quad (9)$$

Recalling the easily proved fact that $E_N \leq E_{N,j}$, we have

$$\langle |\vec{x}_j| \psi | h_j \psi \rangle \leq 0. \quad (10)$$

The claim is that Eq. (10) cannot hold for all j if condition (6) is violated.

First, the term $t_j = \langle |\vec{x}_j| \psi | p_j^2 \psi \rangle$ is positive. To see this, do the $\vec{x}_j$ integration first and note that it then suffices to prove the following for any function, f , of one variable:

$$t = - \int |\vec{x}| f(\vec{x}) \nabla^2 f(\vec{x}) d^3x > 0. \quad (11)$$

Write $g(\vec{x}) = |\vec{x}| f(\vec{x})$ and integrate by parts:

$$\begin{aligned} t &= \int \nabla g(\vec{x}) \cdot \{ |\vec{x}|^{-1} \nabla g(\vec{x}) + g(\vec{x}) \nabla |\vec{x}|^{-1} \} d^3x \\ &= \int \{ |\nabla g(\vec{x})|^2 |\vec{x}|^{-1} - \frac{1}{2} g(\vec{x})^2 \nabla^2 |\vec{x}|^{-1} \} d^3x > 0 \end{aligned}$$

since $\nabla^2 |\vec{x}|^{-1} \leq 0$. (The fact that $g \nabla g = \nabla g^2/2$, together with another integration by parts, was used for the second term.)

The second term in h_j is easy:

$$A_j \equiv q_j \langle |\vec{x}_j| \psi | V(\vec{x}_j) \psi \rangle = q_j z \langle \psi | \psi \rangle = q_j z,$$

with the assumption that ψ is normalized.

The third term is

$$R_j \equiv \int \psi(X) |\vec{x}_j| \sum_{k \neq j} q_j q_k |\vec{x}_j - \vec{x}_k|^{-1} d^{3N}X,$$

where X denotes all the N variables.

If there is binding then Eq. (10) holds for all j and hence, summing over j and using $t_j > 0$, we have that $A \equiv \sum_j A_j > \sum_j R_j \equiv R$. On the one hand,

$A = z \sum_j q_j = zQ$. On the other hand,

$$R = \frac{1}{2} \int \psi(X)^2 \sum_{k \neq j} q_j q_k |\tilde{\mathbf{x}}_j - \tilde{\mathbf{x}}_k|^{-1} (|\mathbf{x}_j| + |\tilde{\mathbf{x}}_k|) d^{3N}X.$$

But $|\tilde{\mathbf{x}}_j| + |\tilde{\mathbf{x}}_k| \geq |\tilde{\mathbf{x}}_j - \tilde{\mathbf{x}}_k|$ (triangle inequality), so that

$$R \geq \frac{1}{2} \sum_{k \neq j} q_j q_k = \frac{1}{2} Q^2 - \frac{1}{2} \sum_j q_j^2.$$

Hence, binding implies that

$$Q^2 - \sum_j q_j^2 < 2zQ,$$

which is precisely Eq. (6). Q.E.D.

This work was partially supported by U. S. National Science Foundation Grant No. PHY-8116101-A01.

¹M. B. Ruskai, Commun. Math. Phys. **82**, 457 (1982).

²I. M. Sigal, Commun. Math. Phys. **85**, 309 (1982).

See also *Mathematical Problems in Theoretical Physics*, Lecture Notes in Theoretical Physics Vol. 153, edited by G. Dell'Antonio, S. Doplicher, and R. Jona-Lasinio (Springer-Verlag, Berlin, 1982), pp. 149-156.

³M. B. Ruskai, Commun. Math. Phys. **85**, 325 (1982).

⁴I. M. Sigal, Institute Mittag Leffler Report No. 12, 1982, revised 1983 (to be published).

⁵R. Benguria and E. H. Lieb, Phys. Rev. Lett. **50**, 1771 (1983).

⁶R. N. Hill, in *Mathematical Problems in Theoretical Physics*, edited by K. Osterwalder, Lecture Notes in Physics Vol. 116 (Springer-Verlag, New York, 1979), pp. 52-56.

⁷G. Zhislin, Tr. Mosk. Mat. Obshch. **9**, 81 (1960).

⁸E. H. Lieb, I. M. Sigal, B. Simon, and W. E. Thirring, "Asymptotic Bulk Neutrality of Large- Z Ions" (to be published), and to be published.

⁹E. H. Lieb, "Bound on the Maximum Negative Ionization of Atoms and Molecules," Phys. Rev. A (to be published).

Bound on the maximum negative ionization of atoms and molecules

Elliott H. Lieb

Departments of Mathematics and Physics, Princeton University, P.O. Box 708, Princeton, New Jersey 08544

(Received 14 November 1983)

It is proved that N_c , the number of negative particles that can be bound to an atom of nuclear charge z , satisfies $N_c < 2z + 1$. For a molecule of K atoms, $N_c < 2Z + K$ where Z is the total nuclear charge. As an example, for hydrogen $N_c = 2$, and thus H^{--} is not stable, which is a result not proved before. The bound particles can be a mixture of different species, e.g., electrons and π mesons; statistics plays no role. The theorem is proved in the static-nucleus approximation, but if the nuclei are dynamical, a related, weaker result is obtained. The kinetic energy operator for the particles can be either $[p - eA(x)/c]^2/2m$ (nonrelativistic with magnetic field) or $\{[pc - eA(x)]^2 + m^2c^4\}^{1/2} - mc^2$ (relativistic with magnetic field). This result is not only stronger than that obtained before, but the proof (at least in the atomic case) is simple enough to be given in an elementary quantum-mechanics course.

I. INTRODUCTION

One of the nonperiodic facts about the Periodic Table is that the number of electrons that can be bound to an atomic nucleus of charge ze is at most $z + 1$, at least as far as present confirmed experimental data go. The theoretical proof of this fact, starting with the Schrödinger equation, is a challenge that has drawn the attention of several authors in recent years.¹⁻¹³

The problem can obviously be extended in two ways: (i) The electrons, which are fermions, can be replaced by bosons or, more generally, by a mixture of particles of different species. (Because of spin, the two-electron problem is well known to be the same as the two-boson problem.) (ii) Instead of a single atom, a molecule can be considered. That these problems are difficult is shown by the fact that it was only recently proved by Ruskai¹ (for bosons) and later Sigal² and Ruskai⁴ (for fermions) that the number of bound particles is not infinite.

The following is a summary of rigorous results to date. Our notation for the maximum particle number is N_c .

(1) Ruskai¹ proved that $N_c \leq (\text{const})z^2$ for bosons. Recently, Sigal³ proved for bosons that for every $\epsilon > 0$ there is a constant C_ϵ such that $N_c \leq C_\epsilon z^{1+\epsilon}$.

(2) For fermions, Sigal² proved that $N_c \leq cz$ with c being some constant. Ruskai⁴ proved that $N_c \leq (\text{const})z^{6/5}$. Sigal³ improved his result to $N_c \leq \alpha(z)z$ with $\alpha(z) \leq 12$ and $\alpha(z) \rightarrow 2$ as $z \rightarrow \infty$.

(3) For bosons, Benguria and Lieb⁵ proved that $N_c \geq \beta(z)z$ with $\beta(z) \rightarrow 1 + \gamma$ as $z \rightarrow \infty$. Here, γ is some number satisfying $0 < \gamma < 1$ and is obtained by solving a Hartree equation. This equation was subsequently solved numerically by Baumgartner^{6,11} with the result $\gamma = 0.21$. Thus, bosons strongly violate the $z + 1$ rule; the Pauli exclusion principle plays a key role in the electron problem.

(4) In a related development, Benguria and Lieb⁷ studied the Thomas-Fermi-von Weizsäcker (TFW) equation—a well-known density-functional equation which is supposed to imitate the Schrödinger equation for fermions. They proved that $N_c \leq z + 0.73$ in the TFW model of an atom. (In TFW theory, the electronic charge

is not quantized.) Thus, on the one hand, TFW theory really imitates the Schrödinger equation and, on the other hand, the TFW result supports the conjecture that $N_c \leq z + \text{const}$ for the Schrödinger equation for arbitrarily large z . Earlier,¹⁴ it had been shown that $Z < N_c < 2Z$ in TFW theory, even for a molecule. Here Ze is the total nuclear charge,

$$Z = \sum_{j=1}^K z_j, \quad (1.1)$$

for a molecule of K atoms with nuclear charges ez_j . Note that in Thomas-Fermi (TF) theory one always has $N_c = Z$ for any molecule.¹⁴

(5) Lieb, Sigal, Simon, and Thirring,⁸ using Sigal's method,³ have proved that $N_c/z \rightarrow 1$ as $z \rightarrow \infty$ in the fermion case.

(6) For $z = 1$ (hydrogen), Hill⁹ proved that three electrons cannot be bound in a quartet ($S = \frac{3}{2}$) state. There does not seem to be any proof of nonbinding for the doublet ($S = \frac{1}{2}$) state with $N = 3$.

All of the above results are for a single atom. The Ruskai and Sigal methods can be extended to the molecular case; this was explicitly done for bosons.¹ The Benguria-Lieb result⁷ in (4) extends to a molecule: $N_c < Z + 0.73$ K. Furthermore, all the results apply to the fixed-nucleus (sometimes called Born-Oppenheimer) approximation.

(7) Zhislin¹⁰ proved that $N_c \geq z$ for an arbitrary mixture of particles (with any statistics) and including nuclear motion. This result extends to a molecule, $N_c \geq Z$.

In this paper it will be proved (Theorem 1) that

$$N_c < 2z + 1 \quad (1.2)$$

for a fixed-nucleus atom and with any mixture of bound particles (with possibly different masses and statistics). Equation (1.2) holds if all the charges are $-e$, but a similar result holds with nonconstant but negative charges [see Eq. (2.11)]. If z is an integer, as in the physical case, (1.2) implies

$$N_c \leq 2z. \quad (1.3)$$

This completes the story for hydrogen, i.e., $N_c = 2$, since it is well known that two electrons can indeed be bound [see (6) above]. H^- is *not* stable. (Incidentally, I am not aware of any proof of the obvious assertion that if N electrons cannot be bound then $M > N$ electrons cannot be bound. Thus, even if the $S = \frac{1}{2}$ case in (6) above were settled, it would not immediately follow from this that $N_c = 2$.) Equation (1.3) also states that two π^- mesons and a muon cannot be bound with $z=1$. Equation (1.2) implies that the critical z to bind two particles is at least 0.5. The exact value¹² is 0.9112.

For large z , Eq. (1.3) is hardly optimal in view of the conjecture that $N_c - z$ is of order unity. For bosons, however, Eq. (1.3) gives the right order of magnitude since $N_c \geq 1.2z$ for large z [see (3) above].

In the case of a molecule of K atoms it will be proved that

$$N_c < 2Z + K \quad (1.4)$$

for fixed nuclei. Thus, for example, the hydrogen molecule cannot bind more than five particles. Again, this holds for arbitrary negative particles with common charge $-e$. For nonconstant charges see Eq. (2.10). A summary of the results of this paper is in Ref. 13.

As stated above, Theorem 1 does not require that the nuclear or the negative particle charges be integral. This generality may be relevant physically because, as pointed out to me by W. Thirring (private communication), particles in solids such as semiconductors may have nonintegral effective charges because of dielectric effects.

A remark should be made about the meaning of "fixed nuclei" in the molecular case. There are two possible interpretations.

Case A: The nuclei, of charges $z_1, \dots, z_K > 0$, have coordinates $R_1, \dots, R_K$ which are arbitrary but which are fixed once and for all, independent of the particle number N .

Case B: For each particle number, the nuclear coordinates R_j are adjusted to minimize the total energy, namely,

$$\hat{E}_N \equiv \inf_{\underline{R}} E_N(\underline{R}) + U(\underline{R}), \quad (1.5)$$

where $\underline{R}$ denotes $\{R_1, \dots, R_K\}$ and $E_N(\underline{R})$ is the electronic ground-state energy depending on $\underline{R}$, and $U(\underline{R})$ is the internuclear interaction. Usually $U(\underline{R})$ is the Coulomb energy

$$e^2 \sum_{1 \leq i < j \leq K} z_i z_j |R_i - R_j|^{-1},$$

but for our purposes $U(\underline{R})$ can be anything.

The result in Eq. (1.4), and more generally in Theorem 1, holds for both Case A and Case B. Case B is, of course, more physical. However, Theorem 1 in Case A implies Theorem 1 in Case B; this is obvious since if binding does not occur for every choice of $\underline{R}$ it certainly does not occur for the minimizing $\underline{R}$. It is Case A (fixed nuclei) that we shall actually consider henceforth.

The case of truly dynamical nuclei cannot be treated as definitively by the method presented here. Nevertheless, some information about this system is contained in Sec.

VI. Section VI also contains the extension of the previously cited results to the case of smeared, but spherical, nuclear charge densities. These results are also shown to hold in the Hartree-Fock theory.

Finally, it will also be proved that Eqs. (1.2) and (1.4) hold if the particle kinetic energy operator, which is

$$p^2/2m = -(\hbar^2/2m)\Delta \quad (1.6)$$

in the usual Schrödinger equation, is replaced by the relativistic expression

$$(p^2 c^2 + m^2 c^4)^{1/2} - mc^2 = (-c^2 \hbar^2 \Delta + m^2 c^4)^{1/2} - mc^2. \quad (1.7)$$

Another variation for which (1.2) and (1.4) hold is the inclusion of a magnetic field in either Eq. (1.6) or (1.7):

$$p^2 \rightarrow [\vec{p} - e\vec{A}(x)/c]^2, \quad (1.8)$$

where $\vec{A}$ is a bounded vector potential. One can even have different $\vec{A}$'s for different particles. Other generalizations of the kinetic energy are also possible, as well as generalizations to potentials other than the Coulomb potential (see the remark in Appendix A).

Not only are the results proved here stronger than that obtained previously [except for (5) above], but the proof itself is much simpler. For the atomic case with one species of fermions or bosons the proof is so short that it can be given in an elementary quantum-mechanics course. In order to display the basic idea as clearly as possible, this atomic case will be treated first in Sec. IV. The method of proof borrows heavily from the proof in Ref. 14, Theorem 7.23 of $N_c < 2Z$ for the TFW theory. (In Ref. 14, N_c was called λ_c .) As mentioned there, the basic idea for the TFW proof in the atomic case is due to Benguria.

II. PRELIMINARIES AND NOTATION

The Hamiltonian for N particles is

$$H_N = \sum_{i=1}^N [T_i - q_i V(x_i)] + \sum_{1 \leq i < j \leq N} q_i q_j |x_i - x_j|^{-1}, \quad (2.1)$$

where

$$V(x) = \sum_{j=1}^K z_j |x - R_j|^{-1} \quad (2.2)$$

is the electric potential produced by K fixed nuclei of charges

$$\underline{z} = \{z_1, \dots, z_K\}$$

located at $\underline{R} = \{R_1, \dots, R_K\}$ with $R_j \in \mathbb{R}^3$. Units are used in which the electron charge e is unity, and $\hbar = 1$. We assume $z_i > 0$, all $i = 1, \dots, K$. The number $-q_i$ is the charge of the i th particle, and we assume $q_i \geq 0$ for $i = 1, \dots, N$.

The operator T_i is the kinetic energy operator for the i th particle and it is assumed to have one of the following three forms:

$$T_i^{(1)} = -\Delta_i/2m_i, \quad (2.3)$$

$$T_i^{(2)} = [i \vec{\nabla}_i + \vec{A}_i(x)]^2 / 2m_i, \quad (2.4)$$

$$T_i^{(3)} = c \{ [i \vec{\nabla}_i + \vec{A}_i(x)]^2 + m_i^2 c^2 \}^{1/2} - m_i c^2. \quad (2.5)$$

Here, $m_i > 0$ is the mass of the i th particle, $c \vec{A}_i(x)/q_i$ is the vector potential applied to the i th particle, and c is the speed of light. $\vec{A}_i(x)$ is assumed to be bounded and to go to zero as $|x| \rightarrow \infty$. Equation (2.3) is obviously a special case of (2.4) but, for simplicity, it is treated separately. It must be noted, however, that if the form $T_i^{(3)}$ is used even for just one particle (for example, i), then every nuclear charge z_j must satisfy $\alpha q_i z_j \leq 2/\pi$, where $\alpha = e^2/\hbar c$ is the fine-structure constant. The reason is that the single-particle operator $T^{(3)} - e^2 q V(x)$ is bounded below (as a quadratic form) if and only if $e^2 q z_j \leq 2/\pi$ for every j . The situation is not changed by the addition of the third term in Eq. (2.1). (See Ref. 15.)

The i th particle can have any one of the three forms of T_i , independently. The m_i , q_i , and $\vec{A}_i$ need not be related for different i . There is one proviso, however. If several particles are of the same type (bosons or fermions) then that group must, of course, have the same m_i , q_i , and T_i . Spin can be included in the usual way. H_N is spin independent. The easiest way to treat spin is to think of the spin coordinate as merely labeling a particle type. Thus, for spin- $\frac{1}{2}$ electrons, there are two kinds of fermions: those with spin up and those with spin down.

The ground-state energy, is defined by

$$E_N = \inf \text{Spec}(H_N) \quad (2.6)$$

$$= \inf_{\psi} \frac{\langle \psi | H_N | \psi \rangle}{\langle \psi | \psi \rangle}, \quad (2.7)$$

the latter being the variational principle. The admissible ψ in Eq. (2.7) must, of course, satisfy the required statistics for the various groups of particles. E_N need not be an eigenvalue, and it will not be if the system is not bound. When a particle is added to the system it can always be placed at infinity (i.e., arbitrarily far away and with arbitrarily small kinetic and potential energy), and we thus have the relation

$$E_{N,j} \geq E_N, \quad j = 1, \dots, N, \quad (2.8)$$

where N, j denotes the $(N-1)$ -particle system with the j th particle removed.

Definition: The N -particle system is said to be *bound* if and only if E_N is an eigenvalue, i.e.,

$$H_N \psi = E_N \psi \quad (2.9)$$

for some $\psi \in L^2(\mathbb{R}^{3N})$.

In this definition it is not required that $E_{N,j}$ be an eigenvalue for any j or that $E_N < E_{N,j}$ for any j . If $E_N < E_{N,j}$ for some j then E_N is automatically an eigenvalue. The word "bound" is *not applicable* to a system that merely has an eigenvalue in the continuum (i.e., E_N is not an eigenvalue but H_N has an eigenvalue greater than $E_{N,j}$ for some j); such a system would only be metastable. Our main result is the following.

Theorem 1: Let $Z = \sum_{j=1}^K z_j$ as before. Let $Q = \sum_{i=1}^N q_i$ be the negative of the total particle charge and let q be the maximum of the q_i . Then, if binding

occurs, the following must be satisfied:

$$Q < 2Z + qK. \quad (2.10)$$

In the atomic case ($K=1$), Eq. (2.10) can be replaced by the slightly stronger requirement

$$Q < 2Z + \sum_{i=1}^N q_i^2 / Q. \quad (2.11)$$

III. STRATEGY OF THE PROOF

Given Eq. (2.9) select one variable, for example, j , and a positive function of one variable, denoted by $1/\phi(x)$, to be determined appropriately. Multiply Eq. (2.9) by

$$\psi^*(x_1, \dots, x_N) / \phi(x_j)$$

and integrate over all the variables, and then take the real part. On the left-hand side there will be four terms:

$$h_j = \langle [\psi / \phi(x_j)] | H_{N,j} | \psi \rangle, \quad (3.1)$$

$$t_j = \text{Re} \langle [\psi / \phi(x_j)] | T_j | \psi \rangle, \quad (3.2)$$

$$-a_j = -q_j \langle [\psi / \phi(x_j)] | V(x_j) | \psi \rangle, \quad (3.3)$$

$$r_j = \left\langle [\psi / \phi(x_j)] \left| \sum_{i \neq j} q_i q_j |x_i - x_j|^{-1} \right| \psi \right\rangle. \quad (3.4)$$

[The right-hand sides of Eqs. (3.1), (3.3), and (3.4) are automatically real.] On the right-hand side there will be $E_N I_j$ with

$$I_j = \langle [\psi / \phi(x_j)] | \psi \rangle. \quad (3.5)$$

Let us assume, provisionally, that all these terms are finite.

Because of Eq. (2.8)

$$E_N I_j \leq h_j. \quad (3.6)$$

To see this, let X_j denote the variables $x_1, \dots, x_N$ with x_j excluded. Consider

$$P_j(X_j, X'_j) = \int \psi(x_j, X_j) [\psi^*(x_j, X'_j) / \phi(x_j)] d^3 x_j, \quad (3.7)$$

so that

$$\Gamma_j(X_j, X'_j) \equiv P_j(X_j, X'_j) / I_j$$

is a properly normalized density matrix. (The positivity of ϕ is crucial here.) Moreover, Γ_j satisfies the correct statistics in the X_j variables (since ψ does) and thus, by the variational principle,

$$h_{N,j} / I_j = \text{Tr}(H_{N,j} \Gamma_j) \geq E_{N,j}. \quad (3.8)$$

This, together with Eq. (2.8), proves (3.6). Thus, binding cannot occur if

$$t_j - a_j + r_j > 0. \quad (3.9)$$

Let $\phi(x)$ be any function of the form

$$\phi(x) = \int |x - y|^{-1} d\mu(y) + C, \quad (3.10)$$

where $d\mu$ is a (positive) measure with $0 < \int d\mu < \infty$ and $C \geq 0$ is a constant. In our application we shall choose ϕ

to be of the same form as V but with different coefficients $\mu_j > 0$:

$$\phi(x) = \sum_{j=1}^k \mu_j |x - R_j|^{-1}. \quad (3.11)$$

It will be proved that for every j

$$t_j > 0 \quad (3.12)$$

for any ψ and ϕ and choice of T_j [see Eqs. (2.3)–(2.5)]. Consequently, binding cannot occur if, for any $j=1, \dots, N$,

$$r_j \geq a_j. \quad (3.13)$$

Actually, we shall not prove Eq. (3.13) for any particular j , but shall prove instead that when Eq. (2.10) or (2.11) is violated then

$$R \equiv \sum_{j=1}^N r_j \geq A \equiv \sum_{j=1}^N a_j. \quad (3.14)$$

[In fact, we shall actually prove that $R > A$ and, therefore, that strict inequality is not really needed in Eq. (3.12). The reason that $R > A$ is given in the second part of Appendix A.] Clearly, Eq. (3.14) implies that Eq. (3.13) holds for some j , and thus Theorem 1 will be proved.

To prove (3.12) we note that the X_j integration in (3.2) can be done after the x_j integration. Therefore, (3.12) will be true if it holds for each X_j , i.e., if it holds for any function of the one variable x_j . Thus, if $f(x) \in L^2(R^3)$ we want to prove that

$$\begin{aligned} 2mt &= \text{Re} \int \bar{\nabla} g^*(x) \cdot \bar{\nabla} [\phi(x)g(x)] d^3x \\ &= \text{Re} \int \bar{\nabla} g^*(x) \cdot [\phi(x) \bar{\nabla} g(x) + g(x) \bar{\nabla} \phi(x)] d^3x \\ &= \int \phi(x) |\bar{\nabla} g(x)|^2 d^3x + \frac{1}{2} \int \bar{\nabla} \phi(x) \cdot [g(x) \bar{\nabla} g^*(x) + g^*(x) \bar{\nabla} g(x)] d^3x \\ &= \int \phi(x) |\bar{\nabla} g(x)|^2 d^3x + \frac{1}{2} \int \bar{\nabla} \phi(x) \cdot \bar{\nabla} [|g(x)|^2] d^3x \\ &= \int \phi(x) |\bar{\nabla} g(x)|^2 d^3x - \frac{1}{2} \int |g(x)|^2 \Delta \phi(x) d^3x > 0. \end{aligned} \quad (3.17)$$

[Note that we have greater than 0 instead of greater than or equal to 0 because $\int \phi |\bar{\nabla} g|^2 > 0$ since $\phi(x) > 0$ for all x ; if $\bar{\nabla} g(x) \equiv 0$ then $f = (\text{const})\phi$, but $\phi \notin L^2$ and $f \in L^2$.]

In the rest of this paper (except Appendix A) we shall assume that (3.15) holds and shall concentrate on proving that Eq. (3.14) holds. As mentioned before, all r_j and a_j are necessarily real and finite.

IV. ATOMS WITH IDENTICAL PARTICLES

Take the nuclear coordinate $R_1=0$, $z_1 \equiv z$ and let $\phi(x) = 1/|x|$. Assume that the particles are identical (bosons or fermions) so that $q_j = q$, $r_j = r$, and $a_j = a$ are independent of j . (Note: If the particles are fermions they are allowed to have spin, in which case $\int dx$ in the following should be understood as $\sum_\sigma \int dx$, where σ is the spin variable.)

We denote $\{x_1, \dots, x_N\}$ by X and assume that ψ is normalized. Taking $j=1$ we have

$$t \equiv \text{Re} \int f^*(x) [(Tf)(x)/\phi(x)] d^3x > 0. \quad (3.15)$$

When $T = p^2/2m$, (3.15) was first proved for f spherically symmetric and real by Benguria, and then for f real by Lieb. This was given in Ref. 14, Lemma 7.21. Baumgartner¹¹ found a more direct proof and also extended (3.15) to complex f . Baumgartner's proof easily extends to $T^{(2)}$, but the proof for $T^{(3)}$ is very different and is given in Appendix A.

In Appendix A a proof of (3.15) under carefully stated conditions on f is given. The following technical point, which is also discussed in Appendix A, has to be considered: We assumed that all the quantities in Eqs. (3.1)–(3.5) are finite. By the condition stated after Eq. (3.10), r_j and a_j are automatically finite. Conceivably, h_j , t_j , and I_j could be infinite. If so, this can be remedied by replacing $\phi(x)$ by $\phi(x) + C$ and then letting $C \rightarrow 0$ at the end. This procedure is also discussed in Appendix A.

Since one of our stated goals is to present a proof of Theorem 1 in the atomic case that is simple enough to be given in an elementary quantum-mechanics course, let us temporarily suspend any reservations about technicalities and give the following proof of (3.15), following Baumgartner's method,¹¹ when $T = p^2/2m$.

The key fact is that $\Delta\phi \leq 0$. Given f and ϕ , define $g(x) = f(x)/\phi(x)$. We then require that

$$2mt = -\text{Re} \int g^*(x) \Delta[g(x)\phi(x)] d^3x > 0. \quad (3.16)$$

By partial integration,

$$a = zq \int [|\psi(X)|^2 |x_1| / |x_1|] d^{3N}X = zq, \quad (4.1)$$

$$r = q^2 \sum_{j=2}^N \int |\psi(X)|^2 |x_1| |x_1 - x_j|^{-1} d^{3N}X \quad (4.2)$$

$$= q^2(N-1) \int |\psi(X)|^2 |x_1| |x_1 - x_2|^{-1} d^{3N}X. \quad (4.3)$$

In going from Eq. (4.2) to (4.3) the fact that $|\psi|^2$ is symmetric was used. This symmetry also implies that the integral in Eq. (4.3) is not changed if $|x_1|$ is replaced by $|x_2|$. Thus,

$$\begin{aligned} r &= \frac{1}{2} q^2 (N-1) \int |\psi(X)|^2 [|x_1| + |x_2|] \\ &\quad \times |x_1 - x_2|^{-1} d^{3N}X \geq \frac{1}{2} q^2 (N-1) \end{aligned} \quad (4.4)$$

since

$$|x_1| + |x_2| \geq |x_1 - x_2|$$

by the triangle inequality.

Therefore,

$$r - a \geq \frac{1}{2} q^2 (N-1) - zq \quad (4.5)$$

and, by Eq. (3.13), binding will not occur if $Q \geq 2z + q$ with $Q = Nq$. This proves Theorem 1 in this special case.

V. GENERAL CASE

Our goal here is to prove (3.14) if (2.11) (atomic case) or (2.10) (molecular case) is violated. In the atomic case without particle symmetry, we cannot use symmetry to go from Eq. (4.2) to (4.3). In the molecular case, even if particle symmetry exists, the obvious choice $\phi(x) = V(x)$ will not work and we shall have to use ϕ as given in Eq. (3.11).

In the atomic case

$$\phi(x) = 1/|x| = V(x)/z$$

does work. Assume ψ to be normalized. For each j , Eqs. (4.1) and (4.2) become

$$a_j = zq_j, \quad (5.1)$$

$$r_j = \sum_{i \neq j} q_i q_j \int |\psi(X)|^2 |x_j| |x_i - x_j|^{-1} d^{3N}X. \quad (5.2)$$

Now sum these over j to obtain [recalling Eq. (3.14)]

$$R - A = \frac{1}{2} \sum_{i,j} q_i q_j \int |\psi(X)|^2 (|x_i| + |x_j|) \times |x_i - x_j|^{-1} d^{3N}X - Qz \quad (5.3)$$

$$\geq \frac{1}{2} \sum_{i,j} q_i q_j - Qz = \frac{1}{2} Q^2 - \frac{1}{2} \sum_i q_i^2 - Qz. \quad (5.4)$$

Again, the triangle inequality $|x| + |y| \geq |x - y|$ has been used. Clearly, if Eq. (2.11) is violated then $R \geq A$ and this proves Theorem 1 in the atomic case.

In the molecular case the following is needed. It does not depend on the dimension d being 3.

Lemma 1: Let $\psi(X)$ be any normalized function in $L^2(\mathbb{R}^{Nd})$ (without any particular symmetry). For $j = 1, \dots, N$ let

$$\phi(x)^{-1} + \phi(y)^{-1} = [\phi(x)\phi(y)]^{-1} [\phi(x) + \phi(y)] = \sum_{j=1}^K \mu_j [\phi(x)|x - R_j|]^{-1} [\phi(y)|y - R_j|]^{-1} (|x - R_j| + |y - R_j|). \quad (5.12)$$

Again, noting that

$$|x - R_j| + |y - R_j| \geq |x - y|$$

we have, upon inserting Eq. (5.12) in (5.11),

$$R \geq \sum_{s=1}^K \mu_s \int |\psi(X)|^2 \sum_{1 \leq i < j \leq N} g_i^s(x_i) g_j^s(x_j) d^{3N}X, \quad (5.13)$$

with

$$\rho_j(x) = \int |\psi(x, X_j)|^2 d^{Nd-d}X_j \quad (5.5)$$

be the one-particle density for particle j . Let $g_1(x), \dots, g_N(x)$ be any given functions of one variable such that $g_j(x_j)\psi(X)$ is in $L^2(\mathbb{R}^{Nd})$ and define

$$\langle g_j \rangle = \int \rho_j(x) g_j(x) d^d x = \int |\psi(X)|^2 g_j(x_j) d^{Nd}X. \quad (5.6)$$

Then

$$\begin{aligned} \text{Re} \int |\psi(X)|^2 \sum_{1 \leq i < j \leq N} g_i^*(x_i) g_j(x_j) d^{Nd}X \\ \geq \frac{1}{2} \left| \sum_{j=1}^N \langle g_j \rangle \right|^2 - \frac{1}{2} \sum_{j=1}^N \langle |g_j|^2 \rangle. \end{aligned} \quad (5.7)$$

This is proved simply by noting that

$$\text{Re} \sum_{i,j} g_i^* g_j = \frac{1}{2} \left| \sum_i g_i \right|^2 - \frac{1}{2} \sum_i |g_i|^2$$

and then using the Schwarz inequality on the first term.

To apply Lemma 1 to the general case, let

$$\tilde{\rho}(x) \equiv \sum_{j=1}^N q_j \rho_j(x) \quad (5.8)$$

be the negative of the single particle charge density for ψ [see Eq. (5.5)]. Let $\phi(x)$ be the potential in Eq. (3.11). We then have

$$A = \sum_{j=1}^N a_j = \sum_{s=1}^K z_s \gamma_s, \quad (5.9)$$

with

$$\gamma_s = \int \tilde{\rho}(x) |x - R_s|^{-1} \phi(x)^{-1} d^3x. \quad (5.10)$$

$$\begin{aligned} R = \sum_{j=1}^N r_j = \sum_{1 \leq i < j \leq N} q_i q_j \int |\psi(X)|^2 |x_i - x_j|^{-1} \\ \times [\phi(x_i)^{-1} + \phi(x_j)^{-1}] d^{3N}X. \end{aligned} \quad (5.11)$$

Let us write [following an idea in Ref. 11)]

$$g_i^s(x) = q_i [\phi(x)|x - R_s|]^{-1}. \quad (5.14)$$

Using Lemma 1 for each s we have

$$\begin{aligned} R \geq \frac{1}{2} \sum_{s=1}^K \mu_s \left\{ \gamma_s^2 - \sum_{i=1}^N q_i^2 \int \rho_i(x) \phi(x)^{-2} \right. \\ \left. \times |x - R_s|^{-2} d^3x \right\}. \end{aligned} \quad (5.15)$$

However,

$$\mu_s / [|x - R_s| \phi(x)] \leq 1,$$

and $q_i \leq q$ (by definition). Hence,

$$R \geq \frac{1}{2} \sum_{s=1}^N (\mu_s \gamma_s^2 - q \gamma_s). \quad (5.16)$$

We shall have $R \geq A$ if

$$\sum_{s=1}^K \gamma_s \{ \mu_s \gamma_s - q - 2z_s \} \geq 0, \quad (5.17)$$

and our aim is to choose the $\{\mu_s\}$ so that Eq. (5.17) is satisfied when condition (2.10) is violated. (Note that γ_s depends on $\{\mu_s\}$.) To this end, let

$$\delta_s = \mu_s \gamma_s / Q, \quad (5.18)$$

$$\beta_s \equiv (2z_s + q) / (2Z + q\kappa).$$

Note that [see Eqs. (5.8) and (5.10)]

$$\sum_{s=1}^K \delta_s = \sum_{s=1}^K \beta_s = 1. \quad (5.19)$$

Suppose we can choose $\{\mu_s\}$ such that

$$\delta_s = \beta_s, \quad s = 1, \dots, K. \quad (5.20)$$

Then the left-hand side of Eq. (5.17) becomes

$$\sum_{s=1}^K \gamma_s \delta_s [Q - (2Z + q\kappa)],$$

and this is non-negative if condition (2.10) is violated.

Thus, showing that the K equations (5.20) in K unknowns have a solution proves Theorem 1. This is done in Appendix B.

VI. THREE GENERALIZATIONS OF THEOREM 1

A. Smeared nuclei

Suppose that the nuclear charge densities, instead of being points, are smeared into *spherically symmetric* distributions about R_j , namely, $z_j |x - R_j|^{-1}$ in Eq. (2.2) is replaced by

$$V_j(x) = \int |x - y - R_j|^{-1} d\mu_j(y), \quad (6.1)$$

where $d\mu_j$ is a spherically symmetric (positive) measure with $\int d\mu_j(y) = z_j$. Then Theorem 1 continues to hold without modification. The proof is as before—with the same $\phi(x)$. One merely has to note that

$$V_j(x) \leq z_j |x - R_j|^{-1} \quad (6.2)$$

for all x and, hence, a_j , defined by Eq. (3.3), is not greater than it would be for the point nucleus.

B. Dynamical nuclei

Suppose that the K nuclear coordinates $R_1, \dots, R_K$ are dynamical variables and the Hamiltonian is

$$\tilde{H}_N = H_N + T_{\text{nuc}} + U(\underline{R}), \quad (6.3)$$

and

$$T_{\text{nuc}} = \sum_{i=1}^K T_i, \quad (6.4)$$

where H_N is given by Eq. (2.1) and each T_i is one of the operators given in Eqs. (2.3)–(2.5). $U(\underline{R})$ is a potential that depends on the nuclear coordinates, as explained after Eq. (1.5). If $U(\underline{R})$ is translation invariant, the eigenvalue equation

$$\tilde{H}_N \psi = \tilde{E}_N \psi, \quad (6.5)$$

with $\tilde{E}_N = \inf \text{Spec}(\tilde{H}_N)$ and $\psi = \psi(X, \underline{R})$, would not generally be expected to have a solution in $L^2(\mathbb{R}^{3N+3K})$; if there are no magnetic fields present then it certainly would not. With magnetic fields present one cannot simply remove the center of mass motion, and the situation is complicated. To avoid technical complications it will be assumed that $U(\underline{R})$ also contains one-body terms which serve to contain the nuclei, and therefore that Eq. (6.5) indeed has an L^2 solution. Physically, this is no real restriction because the confining potential could, for example, be an infinitely high walled box of arbitrarily large size. Because the *negative* particles are *not confined* it is still true that $\tilde{E}_N \leq \tilde{E}_{N,j}$.

In this case a weakened form of Theorem 1 holds. Let us define $E_{N,j}^\infty$ to be the $\inf \text{Spec}(\tilde{H}_{N,j}^\infty)$, with negative particle j removed as before, but with all the nuclear masses set equal to infinity. Alternatively,

$$E_{N,j}^\infty = \inf_{\underline{R}} E_{N,j}(\underline{R}) + U(\underline{R}), \quad (6.6)$$

where $E_{N,j}(\underline{R})$ is the ground-state energy of $H_{N,j}$ as defined in Eq. (2.7). Clearly,

$$E_{N,j}^\infty < \tilde{E}_{N,j} = \inf \text{Spec}(\tilde{H}_{N,j}). \quad (6.7)$$

Theorem 2: Suppose that the system is bound [i.e., Eq. (6.5) holds] and the binding energy satisfies

$$\tilde{E}_N - \tilde{E}_{N,j} \leq E_{N,j}^\infty - \tilde{E}_{N,j} \quad (6.8)$$

for all $j = 1, \dots, N$. Then Eq. (2.10) holds in the molecular case $K > 1$ and Eq. (2.11) holds in the atomic case $K = 1$.

In the physical situation, the right-hand side of Eq. (6.9) is numerically small, but it is a challenge to eliminate this condition.

To prove Theorem 2 we multiply Eq. (6.5) by

$$\psi^*(X, \underline{R}) / \phi(x_j, \underline{R})$$

and integrate over all the variables. Here $\phi(x, \underline{R})$ is as in the proof of Theorem 1, namely, Eq. (3.11) with *fixed constants* μ_j . [It is important that the μ_j do not depend on $\underline{R}$, otherwise the dependence of ϕ on each R_j would not have the form of Eq. (3.10), and thus the positivity of the integrated T_{nuc} term might be lost.]

Let us write (for any fixed j , $1 \leq j \leq N$)

$$\tilde{H}_N = \tilde{H}_{N,j}^\infty + T_{\text{nuc}} + T_j - q_j V(x_j, \underline{R}) + \sum_{i \neq j} |x_i - x_j|^{-1}. \quad (6.9)$$

The first term, $\tilde{H}_{N,j}^\infty = \tilde{H}_{N,j}^\infty(X_j, \underline{R})$, satisfies (as an operator) $\tilde{H}_{N,j}^\infty \geq E_{N,j}^\infty$. Therefore, moving this term to the right-hand side of Eq. (6.5) we obtain less than or equal to 0 as before [using Eq. (6.8)].

The term involving T_j is positive as before. The T_{nuc} term is positive since $\phi(x_j, \underline{R})$, with all variables except R_i fixed, has the correct form [as given by Eq. (3.10)] as a function of R_i . The third term is (after summing on j) the same as in Eq. (5.9) but with

$$\gamma_s = \int \int |\psi(X, \underline{R})|^2 \sum_{j=1}^N [q_j |x_j - R_s|^{-1} / \phi(x_j, \underline{R})] \times d^3N X d^3K R. \quad (6.10)$$

In the atomic case, $\phi(x, R) = |x - R|^{-1}$ and the proof proceeds exactly as before. In the molecular case, Eq. (5.16) holds with γ_s given by Eq. (6.10). [An obvious modification of Lemma 1 and Eqs. (5.11)–(5.15) is needed to include the $\underline{R}$ dependence; the basic observation is that the Schwarz inequality used in Lemma 1 is still applicable.]

To complete the proof in the molecular case, we require that Eq. (5.20) have a solution with the new definition of γ_s , Eq. (6.10). The proof in Appendix B is easily modified in terms of the appropriate matrix M [which has strictly positive elements and satisfies Eq. (B8)]:

$$M_{sj} = Q^{-1} \int \int |\psi(X, \underline{R})|^2 \times \sum_{k=1}^N q_k |x_k - R_s|^{-1} |x_k - R_j|^{-1} \times \phi(x_k, \underline{R})^{-2} d^3N X d^3K R. \quad (6.11)$$

C. Hartree and Hartree-Fock theories

Theorem 1 applies to the Schrödinger equation, but the conclusion remains true in both the Hartree and Hartree-Fock (HF) (either restricted or unrestricted) approximations. Here, the proof for unrestricted HF theory will be given; the proof for the Hartree theory is very similar. It will also be assumed that all the N charges, masses, and kinetic energy operators are identical. We take $q_i = 1$.

In the HF theory, the ground-state energy is defined as in Eq. (2.7), but with ψ restricted to the class of determinantal functions:

$$\psi = (N!)^{-1/2} \det u_i(x_j, \sigma_j) \quad (6.12)$$

and the u_i are orthonormal. As in the earlier definition, the N -particle system is said to be bound if and only if there is a ψ that actually minimizes the energy expression (2.7). As before, $E_N \leq E_{N,j}$. The reader is referred to Ref. 16 for details; in particular, for the proof that binding occurs if $N < Z + 1$ (when $T = T^{(1)}$).

If there is a minimum, the u_i (after possibly an $N \times N$ unitary transformation) satisfy the N coupled HF equations:

$$h u_i = \epsilon_i u_i, \quad (6.13)$$

where h is the single-particle operator

$$h = T - V(x) + U - K. \quad (6.14)$$

Here, T is one of the operators in Eqs. (2.3)–(2.5), U is the direct part of the Coulomb repulsion, and K is the ex-

change part.

Since ψ minimizes the energy, each $\epsilon_i \leq 0$ in Eq. (6.13). The reason is that the dependence of the numerator in Eq. (2.7) on any one u_i is constant plus quadratic, and the latter term is just ϵ_i . If $\epsilon_1 > 0$, for example, then it is easy to see that the $(N-1)$ -particle ψ composed of $u_2, \dots, u_N$ would have an energy strictly below E_N , and this contradicts $E_{N,1} \geq E_N$.

Let $\phi(x)$ be as in Eq. (3.11), multiply Eq. (6.13) by $u_i^*(x)/\phi(x)$, integrate over x , and sum over spins. Then sum over i . The right-hand side is nonpositive. On the left-hand side, the terms involving T are positive, as before. The $V(x)$ term is

$$A = N \int |\psi(X)|^2 V(x_1) / \phi(x_1) d^3N X.$$

The repulsion term, $U - K$, is

$$R = \frac{1}{2} N(N-1) \int |\psi(X)|^2 |x_1 - x_2|^{-1} \times [\phi(x_1)^{-1} + \phi(x_2)^{-1}] d^3N X.$$

(A summation on spins is understood in these two integrals.) The rest of the analysis proceeds as before.

ACKNOWLEDGMENTS

This work was partially supported by U.S. National Science Foundation Grant No. PHY-81-16101-A01.

APPENDIX A: KINETIC-ENERGY INEQUALITY

1. Proof of Eq. (3.15)

Our first goal is to prove Eq. (3.15) when T is one of the three forms in Eqs. (2.3)–(2.5). The potential ϕ is defined by

$$\phi(x) = \int |x - y|^{-1} d\mu(y) + C, \quad (A1)$$

with $d\mu$ a positive measure, $0 < \int d\mu < \infty$, and $C \geq 0$. The function f satisfies

$$f \text{ and } Tf \in L^2(R^3). \quad (A2)$$

Note that if $C > 0$ then $1/\phi(x)$ is a bounded function and $f(x)/\phi(x)$ is automatically in L^2 .

The following Lemma 2 validates the assertions used in Sec. III provided

$$\psi(X)/\phi(x_j) \in L^2(R^{3N}).$$

Subsection II below shows how to deal with the case $\psi/\phi \notin L^2$ by taking $C > 0$ and then letting $C \rightarrow 0$. [Technical remark: In the nonrelativistic case a solution to Eq. (2.9) automatically satisfies $T_j \psi \in L^2(R^{3N})$ and hence $Tf \in L^2(R^3)$. The analogous statement is known to hold in the relativistic case if $e^2 q_i z_i < \frac{1}{2}$ (for all i), which is less than the critical value $2/\pi$. Thus, there is possibly a minor technical gap in the relativistic case.] Lemma 2 in the relativistic case was originally proved only for $\bar{A}(x) \equiv 0$. The extension to $\bar{A}(x) \neq 0$ follows from Lemma 3, which is a joint work with Michael Loss. Lemma 3 shows that Lemma 2 (relativistic) follows automatically from Lemma 2 (nonrelativistic). Nevertheless, the original proof of Lemma 2 for $T^{(3)}$ with $\bar{A}(x) \equiv 0$ is not

without interest and is therefore given below.

Lemma 2: Assume Eqs. (A1), (A2), and

$$f(x)/\phi(x) \in L^2(\mathbb{R}^3).$$

Let T be any one of the operators in Eqs. (2.3)–(2.5), with $\bar{A}(x)$ bounded in Eqs. (2.4) and (2.5). Then

$$h(x) \equiv f^*(x)(Tf)(x)/\phi(x) \quad (\text{A3})$$

is in L^1 and

$$t = \text{Re} \int h(x) d^3x > 0. \quad (\text{A4})$$

Proof: At first assume $C > 0$ so that $1/\phi$ is bounded. By a simple density argument, we can suppose $f \in C_0^\infty$ (infinitely differentiable functions of compact support). We can also assume [by replacing $d\mu$ by $\exp(-\epsilon x^2) * d\mu$, for example, and then using dominated convergence] that ϕ and $1/\phi \in C^\infty \cap L^\infty$. Unfortunately, the limiting argument just cited will only allow us to conclude that $t \geq 0$. The proof that $t > 0$ is given at the very end.

$T^{(1)} = -\Delta$: The proof given in Eqs. (3.16) and (3.17) is completely rigorous for such functions. Another proof is given below in connection with $T^{(3)}[\bar{A}(x) \equiv 0]$.

$T^{(2)} = [i\vec{\nabla} + \bar{A}(x)]^2$: We set $g(x) = f(x)/\phi(x)$ as before and follow the same manipulation as in Eq. (3.17), but replacing $\vec{\nabla}$ by $\vec{\nabla} - i\bar{A}(x)$. The only essential difference is that we get an extra term

$$i \int \vec{\nabla} \phi(x) \cdot \bar{A}(x) |g(x)|^2 d^3x,$$

but this vanishes when the real part is taken. Strict positivity follows from the fact that $(i\vec{\nabla} - \bar{A})g$ cannot vanish identically.

$T^{(3)} = (-\Delta + 1)^{1/2} - 1$: Here we have to work in momentum space. $T^{(3)}f$ is defined to be the function whose Fourier transform $(T^{(3)}f)^\wedge$ is

$$[(k^2 + 1)^{1/2} - 1]\hat{f}(k),$$

with

$$\hat{f}(k) = \int \exp(ik \cdot x) f(x) d^3x.$$

Again, let $g = f/\phi$. Since $(ab)^\wedge = \hat{a} * \hat{b}$, and since $(|x|^{-1})^\wedge = 4\pi/k^2$, we have

$$\begin{aligned} 2\pi^2 t = \text{Re} \int d\mu(y) \int \hat{g}^*(k) |k - q|^{-2} e^{iy \cdot (k - q)} \\ \times m(q) \hat{g}(q) d^3k d^3q \\ + (C/4\pi) \int |g(k)|^2 m(k) d^3k, \end{aligned} \quad (\text{A5})$$

with $m(q) = (q^2 + 1)^{1/2} - 1$. Clearly, the second term on the right-hand side of Eq. (A5) is positive. As for the first term, it is sufficient to prove strict positivity for each y but, since

$$\exp[iy \cdot (k - q)]$$

is a product function, it suffices to prove that the kernel

$$K(k, q) = |k - q|^{-2} [m(k) + m(q)] \quad (\text{A6})$$

is positive definite.

Let us temporarily return to $T^{(1)} = -\Delta$. In this case

$m(k)$ in Eq. (A6) is replaced by k^2 . However,

$$k^2 + q^2 = |k - q|^2 + 2k \cdot q \quad (\text{A7})$$

and thus

$$K(k, q) = 1 + 2k \cdot q |k - q|^{-2}. \quad (\text{A8})$$

The kernel 1 in Eq. (A8) is clearly positive semidefinite. For the second term, note that $|k - q|^{-2}$ is positive definite and $k \cdot q$ is a product function; thus, K is positive definite and we have a second proof for the $T^{(1)}$ case.

Returning to $T^{(3)}$, let us write

$$\begin{aligned} K(k, q) = [n(k) + n(q)]^{-1} |k - q|^{-2} [n(k) + n(q)] \\ \times [m(k) + m(q)], \end{aligned} \quad (\text{A9})$$

with $n(k) = (k^2 + 1)^{1/2}$. The last two factors in Eq. (A9) are

$$B(k, q) = k^2 + q^2 + 2[n(k) - 1][n(q) - 1]. \quad (\text{A10})$$

Now the first factor can be written as an integral over product functions, namely,

$$\int_0^\infty ds \exp\{-s[n(k) + n(q)]\},$$

and this does not affect the positive definiteness. As for the rest we have two terms, namely,

$$|k - q|^{-2} (k^2 + q^2),$$

which is positive definite as we just proved for $T^{(1)}$, and

$$|k - q|^{-2} [n(k) - 1][n(q) - 1].$$

But this last term is a positive definite kernel times a product function, so it too is positive definite. This concludes the proof that $t > 0$ if $C > 0$ and $\bar{A}(x) \equiv 0$ for $T^{(3)}$.

Thus far we have proved that $t > 0$ when (i) $f \in C_0^\infty$, $\phi \in C^\infty$ and (ii) $C > 0$. In all cases we found that $t = Q(g, g)$ with Q being a positive definite quadratic form and $g = f/\phi$. Let us first remove condition (i). There exist sequences f^n and $d\mu^n$ such that $\phi^n \rightarrow \phi$ pointwise almost everywhere and $f^n \rightarrow f$, $Tf^n \rightarrow Tf$, and

$$g^n = f^n/\phi^n \rightarrow g = f/\phi$$

in L^2 . Then

$$t = \lim_{n \rightarrow \infty} t^n = \lim_{n \rightarrow \infty} Q(g^n, g^n) \geq Q(g, g) > 0.$$

Finally, we want to let $C \rightarrow 0$ (if that is the case at hand). With f and ϕ fixed, let $\phi_C = \phi + C$, $g_C = f/(\phi + C)$. With $Y_C = \phi/(\phi + C)$, we have that $0 < Y_C \leq 1$, and $Y_C \rightarrow 1$ pointwise almost everywhere. Then $t_C \rightarrow t$ by dominated convergence. Also, $g_C \rightarrow g = f/\phi$ and $\hat{g}_C \rightarrow \hat{g}$ in L^2 . Again,

$$t = \lim_{C \rightarrow 0} t_C = \lim_{C \rightarrow 0} Q(g_C, g_C) \geq Q(g, g) > 0.$$

This completes the proof for $T^{(1)}$, $T^{(2)}$, and $T^{(3)}[\bar{A}(x) \equiv 0]$. The general case $T^{(3)}[\bar{A}(x) \neq 0]$ follows from the $T^{(2)}$ case and Lemma 3, in which C is the multiplication operator $1/\phi(x)$ and B is the operator $T^{(2)}$, and the integral representation

$$(B+1)^{1/2}-1=\pi^{-1}B\int_1^\infty(x-1)^{1/2}x^{-1}(x+B)^{-1}dx, \quad (\text{A11})$$

Q.E.D.

Lemma 3 (in collaboration with M. Loss): Let H be a Hilbert space with an inner product $(\cdot, \cdot)$ and let B and C be non-negative, self-adjoint linear operators with domain $D(B), D(C)$. Suppose that

$$\begin{aligned} & \text{(i) } (B+x)^{-1}: D(C) \rightarrow D(C), \text{ all } x > 0 \\ & \text{(ii) } \operatorname{Re}(B\phi, C\phi) > 0 \text{ (respectively, } \geq 0), \\ & \quad \text{all } 0 \neq \phi \in D(B) \cap D(C) \\ & \text{(iii) } g(\lambda) = \int d\mu(y)(s\lambda+t)(\lambda+y)^{-1} \end{aligned} \quad (\text{A12})$$

with $s, t \geq 0$, $s+t > 0$, and $\mu \neq 0$ a non-negative Borel measure on R with $\mu\{(-\infty, 0]\} = 0$ and

$$\int d\mu(y)(1+y)^{-1} < \infty.$$

Then

$$\begin{aligned} & \operatorname{Re}(g(B)\phi, C\phi) > 0 \text{ (respectively, } \geq 0), \\ & \quad \text{all } \phi \in D(g(B)) \cap D(C). \end{aligned} \quad (\text{A13})$$

Proof: First, consider the special case

$$g(\lambda) = g_x(\lambda) = (s\lambda+t)(\lambda+x)^{-1}$$

for some fixed $x > 0$. Then $G_x = g_x(B)$ is bounded and we want to prove that $I = (G_x\phi, C\phi)$ satisfies $\operatorname{Re} I > 0$ (respectively, ≥ 0) for all $0 \neq \phi \in D(C)$. Let $\psi = (B+x)^{-1}\phi$, whence

$$0 \neq \psi \in D(B) \cap D(C).$$

Since $(B+x)\psi \in D(C)$ then ψ and $B\psi \in D(C)$. Thus,

$$I = ((sB+t)\psi, C(B+x)\psi) = s(B\psi, CB\psi) + sx(B\psi, C\psi) + t(\psi, CB\psi) + tx(\psi, C\psi).$$

Since $(\psi, CB\psi) = (C\psi, B\psi) = (B\psi, C\psi)^*$, we have

$$\operatorname{Re} I \geq (sx+t)\operatorname{Re}(B\psi, C\psi) > 0 \text{ (respectively, } \geq 0).$$

Now let $I = (g(B)\phi, C\phi)$ with

$$\phi \in D(g(B)) \cap D(C)$$

and, for $0 < \epsilon < 1$,

$$I_\epsilon = (g^\epsilon(B)\phi, C\phi)$$

with

$$g^\epsilon(\lambda) = \int d\mu^\epsilon(y)(s\lambda+t)/(\lambda+y)$$

and where μ^ϵ is μ restricted to the interval $(\epsilon, 1/\epsilon)$. Clearly, $g^\epsilon(B)$ is bounded and

$$I_\epsilon = \int d\nu(\phi, C\phi; \lambda) \int d\mu^\epsilon(y)(s\lambda+t)/(\lambda+y),$$

where $\nu(\phi, \phi'; \cdot)$ is the spectral measure of B associated with ϕ, ϕ' . We want to show that

$$I_\epsilon = \int d\mu^\epsilon(y) \int d\nu(\phi, C\phi; \lambda)(s\lambda+t)/(\lambda+y).$$

Fubini's Theorem cannot be used since ν is not positive. However, by the polarization identity,

$$\begin{aligned} d\nu(\phi, C\phi; \lambda) &= \frac{1}{4} d\nu(a, a; \lambda) - \frac{1}{4} d\nu(b, b; \lambda) \\ &\quad + \frac{1}{4} id\nu(c, c; \lambda) - \frac{1}{4} id\nu(d, d; \lambda), \end{aligned}$$

where $a = \phi + C\phi$, $b = \phi - C\phi$, $c = \phi - iC\phi$, and $d = \phi + iC\phi$. Each of these four measures is non-negative and, since $g^\epsilon(B)$ is bounded, each integral is finite. Thus, we can exchange the order of integration and

$$I_\epsilon = \int d\mu^\epsilon(x) M(x)$$

with

$$M(x) = (G_x\phi, C\phi)$$

and $\operatorname{Re} M(x) > 0$ (respectively, ≥ 0). Therefore, $\lim_{\epsilon \rightarrow 0} \operatorname{Re} I_\epsilon > 0$ (respectively, ≥ 0). On the other hand,

$$I - I_\epsilon = ([g(B) - g^\epsilon(B)]\phi, C\phi).$$

Since $\phi \in D(g(B))$, it is easy to see that $[g(B) - g^\epsilon(B)]\phi \rightarrow 0$. Thus,

$$I = \lim_{\epsilon \rightarrow 0} I_\epsilon.$$

Q.E.D.

Remark: Suppose that $\tilde{g}(\lambda)$ is another function with the same kind of representation as in Eq. (A12). Then, starting with Eq. (A13) and with the pair $g(B), C$ instead of B, C , one can apply Lemma 3 to $\tilde{g}(C)$ and deduce that

$$\begin{aligned} & \operatorname{Re}(g(B)\phi, \tilde{g}(C)\phi) > 0 \text{ (respectively, } \geq 0), \\ & \quad \text{all } \phi \in D(g(B)) \cap D(\tilde{g}(C)). \end{aligned} \quad (\text{A14})$$

It is merely necessary to verify that for all $x > 0$,

$$(C+x)^{-1}: D(g(B)) \rightarrow D(g(B)).$$

This implies the following generalization of the results of this paper:

(i) The relativistic kinetic energy (with magnetic field) can be generalized to any function g of $[\vec{p} - \vec{A}(x)]^2$ that has the form of Eq. (A12).

(ii) The Coulomb potential $1/|x|$ can be replaced (everywhere) in Eq. (2.1) by $v(x) = 1/w(|x|)$ for any function w with the representation

$$w(|x|) = \int d\mu(y)(s|x|)(|x|+y)^{-1}. \quad (\text{A15})$$

With $s > 0$ and $\mu \geq 0$. For example, $1/|x| \rightarrow |x|^{-p}$, $0 < p < 1$ is allowed. It is easy to check that $C = w(|x|)$ satisfies

$$(C+\lambda)^{-1}: D(T^{(i)}) \rightarrow D(T^{(i)}), \quad i=1,2,3$$

for $\lambda > 0$. It is also necessary to check that the "triangle inequality"

$$w(|x|) + w(|z|) \geq w(|x-z|)$$

holds, and this is easily seen to be the case from Eq. (A15). [It is the requirement of the triangle inequality that dictates $s|x|$, instead of $s|x|+t$, in Eq. (A15).]

2. Eliminating infinity

After Eq. (3.5) we made the assumption that all quantities in Eqs. (3.1)–(3.5) were finite. Conceivably, this need not be true with ϕ given by Eq. (3.11). To remedy this defect replace ϕ in Eq. (3.11) by

$$\phi_C(x) = \phi(x) + C, \quad C > 0.$$

Then all quantities are finite. Denote R (respectively, A) with ϕ_C by R_C (respectively, A_C). Binding cannot occur if $R_C > A_C$ for any C (here, the fact that $t > 0$ is ignored).

As $C \rightarrow 0$, R_C and A_C have finite limits R and A which, by dominated convergence, are the R and A given in Eq. (3.14). Thus, it suffices to show that $R > A$ when condition (2.10) or (2.11) is violated. In the earlier proof in Secs. IV and V it was shown that $R \geq A$ by using the triangle inequality

$$|x_i - x_j| \leq |x_i| + |x_j|.$$

This inequality will now be investigated more closely to show that, in fact, $R > A$.

If we look at Eq. (5.11), for example, we have, after integrating over the variables other than $x_i = x$ and $x_j = y$, an expression of the form

$$L \equiv \int f(x, y) (|x| + |y|) |x - y|^{-1} d^3x d^3y, \quad (A16)$$

and we wish to show that

$$L > M \equiv \int f(x, y) d^3x d^3y > 0. \quad (A17)$$

Note that $f(x, y)$ is a non-negative function in L^1 , and not a distribution. The function

$$g(x, y) = (|x| + |y|) |x - y|^{-1-1}$$

satisfies $g \geq 0$ and $g = 0$ if and only if $y = -bx$ with $b \geq 0$. The set on which this occurs has six-dimensional Lebesgue measure zero. Thus, $g > 0$ almost everywhere. Since $f > 0$ on a set of positive measure, $\int fg > 0$ and hence Eq. (A17) holds.

APPENDIX B: SOLUTION OF EQ. (5.20)

Let μ denote $(\mu_1, \dots, \mu_K)$ and consider the function

$$F(\mu) = \sum_{s=1}^K [\delta_s(\mu) - \beta_s]^2 \quad (B1)$$

defined on the positive orthant D : $\mu_i \geq 0$, but excluding the origin $\mu = 0$. The β_s are fixed, strictly positive constants satisfying $\sum_s \beta_s = 1$. δ_s is of the form

$$\delta_s = \int \rho(x) [\mu_s |x - R_s|^{-1} / \phi(x)] d^3x, \quad (B2)$$

with $\int \rho = 1$, and

$$\phi(x) = \sum_s \mu_s |x - R_s|^{-1}.$$

Equation (B2) implies $\sum \delta_s = 1$.

Now $\delta_s(\mu)$ is continuous on D (in particular, $\delta_s(\mu) = 0$ if $\mu_s = 0$) and homogeneous of degree zero, i.e., $\delta_s(\lambda\mu) = \delta_s(\mu)$. Therefore, $F(\mu)$ has a minimum on D . We want to show that this minimum is zero, whence $\delta_s(\mu) = \beta_s$ for all s . Let μ be a minimum point and sup-

pose (without loss of generality) that $\mu_1, \dots, \mu_t > 0$, $\mu_{t+1}, \dots, \mu_K = 0$. (Not all the μ_s can vanish since 0 is not in D .) For $1 \leq s \leq t$, $\delta_s(\mu)$ is differentiable in $\mu_1, \dots, \mu_t$ and for $t+1 \leq s \leq K$, $\delta_s(\mu) \equiv 0$ in a $(t-1)$ -dimensional neighborhood of this point. Therefore, at the minimum,

$$0 = \frac{\partial F}{\partial \mu_j} = 2 \sum_{s=1}^t \mu_s M_{sj} (\delta_j - \beta_j - \delta_s + \beta_s) \quad (B3)$$

for $1 \leq j \leq t$, where M is the t -square matrix

$$M_{sj} = \int \rho(x) |x - R_s|^{-1} |x - R_j|^{-1} \phi(x)^{-2} d^3x. \quad (B4)$$

Clearly, M is symmetric and positive semidefinite and, most importantly, M has positive matrix elements.

Eq. (B3) can be rewritten in the following way (recalling that $\mu_s > 0$ for $1 \leq s \leq t$):

$$Nv = v, \quad (B5)$$

where N is a matrix and v is a vector given by

$$v_s = (\delta_s - \beta_s) \delta_s^{1/2}, \quad (B6)$$

$$N_{sj} = M_{sj} \mu_s \mu_j (\delta_s \delta_j)^{-1/2}. \quad (B7)$$

The fact that

$$\sum_s \mu_s \mu_j M_{sj} = \delta_j \quad (B8)$$

has been used which, in terms of N , reads

$$Nw = w, \quad (B9)$$

with

$$w_s = \delta_s^{1/2}. \quad (B10)$$

Now N is symmetric and has strictly positive matrix elements. By the Perron-Frobenius theorem, N has a unique eigenvalue of largest modulus λ . Moreover, this eigenvalue is positive and has only one eigenvector u , which (up to a phase) has strictly positive components. Equation (B9) implies that $\lambda = 1$ and $u = w$ for, otherwise, taking the inner product of Eq. (B9) with u we would obtain $(\lambda - 1)(u, w) = 0$, which is impossible since $(u, w) > 0$. Thus, the solution to Eq. (B5) is

$$v = cw, \quad (B11)$$

where c is a constant. This means that

$$\delta_s - \beta_s = c$$

for

$$1 \leq s \leq t. \quad (B12)$$

Summing this on s we obtain (since $\delta_s = 0$ for $s > t$)

$$1 - \sum_{s=1}^t \beta_s = ct. \quad (B13)$$

If $t = K$, the left-hand side of Eq. (B13) is zero and we are finished. If $t < K$, then $c > 0$. In the latter case, replace $\mu_s = 0$ by $\mu_s = \epsilon$ for $t < s \leq K$ and with $\epsilon > 0$. It is easy to see that δ_s decreases for $1 \leq s \leq t$ and δ_s becomes strictly positive for $t < s \leq K$. If ϵ is small enough, $(\delta_s - \beta_s)^2$ will decrease for all $1 \leq s \leq K$. Thus, $F(\mu)$ will decrease; as this is a contradiction, $t = K$ and the proof is complete.

- ¹M. B. Ruskai, Commun. Math. Phys. **82**, 457 (1982).
- ²I. M. Sigal, Commun. Math. Phys. **85**, 309 (1982). See also *Mathematical Problems in Theoretical Physics*, Vol. 153 of *Lecture Notes in Theoretical Physics* (Springer, Berlin, 1982), pp. 149–156.
- ³I. M. Sigal, Institute Mittag Leffler Report No. 12 (1982, revised in 1983), Ann. Phys. (to be published).
- ⁴M. B. Ruskai, Commun. Math. Phys. **85**, 325 (1982).
- ⁵R. Benguria and E. H. Lieb, Phys. Rev. Lett. **50**, 1771 (1983).
- ⁶B. Baumgartner, J. Phys. A (to be published).
- ⁷R. Benguria and E. H. Lieb, (unpublished).
- ⁸E. H. Lieb, I. M. Sigal, B. Simon, and W. E. Thirring (unpublished). See also Phys. Rev. Lett. **52**, 994 (1984).
- ⁹R. N. Hill, *Mathematical Problems in Theoretical Physics. Proceedings of the International Conference on Mathematical Physics, Lausanne, 1979*, Vol. 116 of *Lecture Notes in Physics* edited by K. Osterwalder (Springer, New York, 1979), pp. 52–56.
- ¹⁰G. Zhislin, Trudy, Mosk. Mat. Obšč. **9**, 81 (1960).
- ¹¹B. Baumgartner, Lett. Math. Phys. **1**, 439 (1983).
- ¹²F. H. Stillinger and D. K. Stillinger, Phys. Rev. A **10**, 1109 (1974); F. H. Stillinger, J. Chem. Phys. **45**, 3623 (1966).
- ¹³E. H. Lieb, Phys. Rev. Lett. **52**, 315 (1984).
- ¹⁴E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981). Erratum: **54**, 311(E) (1982).
- ¹⁵I. Daubechies and E. H. Lieb, Commun. Math. Phys. **90**, 511 (1983).
- ¹⁶E. H. Lieb and B. Simon, J. Chem. Phys. **61**, 735 (1974); Commun. Math. Phys. **53**, 185 (1977). Theorem 2.4 states, inter alia, that $\epsilon_1, \dots, \epsilon_N$ are the N lowest points of the spectrum of h . This is correct, but the proof is wrong. A correct proof is given in the earlier summary: E. H. Lieb, in *Proceedings of the International Congress of Mathematicians, Vancouver, 1974*, Vol. 2, pp. 383–386.

Approximate Neutrality of Large- Z Ions^{*}

Elliott H. Lieb¹, Israel M. Sigal², Barry Simon³ and Walter Thirring⁴

¹ Departments of Mathematics and Physics, Princeton University, Princeton, NJ 08544, USA

² Department of Mathematics, University of Toronto, Toronto, Canada M5S 1A1

³ Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA

⁴ Institute for Theoretical Physics, University of Vienna, Vienna, Austria

Abstract. Let $N(Z)$ denote the number of electrons which a nucleus of charge Z can bind in non-relativistic quantum mechanics (assuming that electrons are fermions). We prove that $N(Z)/Z \rightarrow 1$ as $Z \rightarrow \infty$.

1. Introduction

This paper is a contribution to the exact study of Coulombic binding energies in quantum mechanics. Let $H(N, Z)$ denote the Hamiltonian

$$H(N, Z) = \sum_{i=1}^N (-\Delta_i - Z|x_i|^{-1}) + \sum_{i < j} |x_i - x_j|^{-1},$$

and let $E(N, Z)$ denote its minimum over all fermion states (we suppose there are two spin states allowed, although any fixed number could be accommodated). For comparison purpose, we let $E_b(N, Z)$ denote the same minimum, but over all states (taken on a totally symmetric wave function, hence b for boson).

It is a fundamental result of Ruskai [9] for bosons, and Sigal [11] for fermions (see also Ruskai [10]) that there exists $N(Z), N_b(Z)$ so that, for all $j = 0, 1, \dots$,

$$E(N(Z), Z) = E(N(Z) + j, Z); \quad E_b(N_b(Z), Z) = E_b(N_b(Z) + j, Z).$$

We let $N(Z)$ (respectively $N_b(Z)$) denote the smallest number for which the first (respectively second) equality holds for all j . Sigal [12] showed that

$$\overline{\lim} [N(Z)/Z] \leq 2, \quad \lim [\ln N_b(Z)/\ln Z] \leq 1, \tag{1.1}$$

and then Lieb [6, 7] proved the bounds

$$N(Z) < 2Z + 1, \quad N_b(Z) < 2Z + 1 \tag{1.2}$$

which implies, in particular, that a doubly ionized hydrogen atom is unstable.

^{*} Research partially supported by the NSERC under Grant NA7901 and by the USNSF under Grants DMS-8416049 and PHY 85-15288-A01

Zhislin [15] proved that $N(Z) \geq Z$ and $N_b(Z) \geq Z$. A more detailed review of the history and status of this and related problems is given in [7].

Our main goal in this paper is to show that

Theorem 1.1. $\lim_{Z \rightarrow \infty} [N(Z)/Z] = 1$.

That is, asymptotically, the excess charge in negative ions is a small fraction of the total charge. While this is physically reasonable, and partially captures the observed fact that in nature there are no highly negative ions, it is not as “obvious” as it might appear at first. For Benguria and Lieb [1] have shown that

$$\lim N_b(Z)/Z > 1.$$

(They actually prove that $\lim$ is at least the critical charge for the Hartree equation which is rigorously known to lie between 1 and 2, and numerically [16] is about 1.2.) Thus, the Pauli principle will enter into our proof of Theorem 1.1.

Part of our argument closely follows that in Sigal [12] (see also Cycon et al. [13]). We differ from Sigal in one critical aspect. He gets a factor of 2 in (1.1) by using the obvious fact that if one has $2Z + 1$ electrons surrounding a nucleus of charge Z , one can always gain classical energy by taking the electron farthest from the nucleus to infinity. We will exploit the fact that if you have $Z(1 + \varepsilon)$ electrons and Z is large, classical energy can be gained by taking *some* electron to infinity. We will prove precisely this fact in Sect. 3. Actually, for technical reasons, we will need a slightly stronger result, also proven there. Unfortunately, our proof of this key classical fact is by contradiction, using a compactness result. Hence our proof is non-constructive, which means that we have no estimates on how large Z has to be for $N(Z)/Z$ to be bounded by $1 + \varepsilon$ for any given ε .

The theorem we prove in Sect. 3 says that, with $N = Z(1 + \varepsilon)$ point electrons, one gains classical energy by taking some electron to infinity. We prove this by appealing to an analogous result for a “fluid” of negative charge: if $\int d\rho(x) \geq Z(1 + \varepsilon)$, then for some x_0 in $\text{supp } \rho$ one has $Z|x_0|^{-1} - \int |x_0 - y|^{-1} d\rho(y) \geq 0$. This fact is proven in Sect. 2. It is interesting that our results in quantum potential theory require various results in classical potential theory.

In Sect. 4, we construct a partition of unity in $\mathbb{R}^{3N}$ —the result of Sect. 3 is needed only to assure that certain sets over $\mathbb{R}^{3N}$. Given this partition, the actual proof of Theorem 1.1. in Sect. 5 follows Sigal [12]. Section 6 provides some additional remarks.

2. Classical Continuum Theorem

Theorem 2.1. *Let ρ be a nonzero finite (positive) measure on $\mathbb{R}^3$ which is not a point mass at 0, and let ϕ_ρ be its potential, i.e.,*

$$\phi_\rho(x) = \int |x - y|^{-1} d\rho(y). \quad (2.1)$$

Then, for any $\varepsilon > 0$, the set of points $x \neq 0$ such that

$$\phi_\rho(x) \geq (1 - \varepsilon)|x|^{-1} \rho(\mathbb{R}^3) \quad (2.2)$$

has positive ρ measure.

Proof. Let T_ε denote the set of points in $\mathbb{R}^3 \setminus \{0\}$ such that (2.2) holds. Our goal is to show that $\rho(T_\varepsilon) > 0$. First, let us eliminate any possible small point mass at 0 by defining $\tilde{\rho} = \rho - c\delta(x)$. For some $0 \leq c < 1$, $\tilde{\rho}(\{0\}) = 0$. Additionally, defining $\tilde{\varepsilon} = \varepsilon\rho(\mathbb{R}^3)/\tilde{\rho}(\mathbb{R}^3)$, one sees that proving the theorem for ε and ρ is equivalent to proving it for $\tilde{\varepsilon}$ and $\tilde{\rho}$ with $(1 - \tilde{\varepsilon})\rho(\mathbb{R}^3) > c$. It suffices to assume, therefore, that $\tilde{\varepsilon} = \varepsilon$, $\tilde{\rho} = \rho$ and $\rho(\{0\}) = 0$, and we shall do so henceforth.

Let B denote the set of $x \neq 0$ for which $\phi_\rho(x) = \infty$. If $\rho(B) > 0$ we are trivially done, so assume $\rho(B) = 0$. Since $\rho(\{0\}) = 0$, this means that ϕ_ρ is finite ρ -a.e. and we can apply Baxter's Theorem 2 [17]. If we define the measure $\mu = (1 - \varepsilon)\{\int d\rho\}\delta(x)$, this theorem asserts the existence of a (positive) measure γ such that

- (a) $\gamma \leq \rho$ and $\gamma(\mathbb{R}^3) \geq \rho(\mathbb{R}^3) - \mu(\mathbb{R}^3) = \varepsilon \int d\rho > 0$,
- (b) $\phi_\rho(x) = \phi_\gamma(x) + \phi_\mu(x)$ γ -a.e..

(In Baxter's notation, $\rho = \nu$, $\rho - \gamma = \lambda$ and $\mu = \mu$.) Thus, $\rho(T_\varepsilon) \geq \gamma(T_\varepsilon) = \gamma(\mathbb{R}^3) > 0$. $\square$

Remark. One can also prove this theorem by appealing to Choquet's theorem [2, 5].

3. Classical-Discrete Theorem

Theorem 3.1. *For any ε , there exists N_0 so that, for all sets $\{\vec{x}_a\}_{a=1}^N$ of $N \geq N_0$ points, we have*

$$\max_b \left\{ \sum_{a \neq b} \frac{1}{|\vec{x}_a - \vec{x}_b|} - \frac{(1 - \varepsilon)N}{|\vec{x}_b|} \right\} \geq 0.$$

Remarks. This is clearly a classical analog of the quantum theorem that we are seeking. It says that if the electron excess over the nuclear charge above Z is more than $\varepsilon(1 - \varepsilon)^{-1}Z$, then one gains energy by moving at least one of the electrons off to infinity.

2. Unfortunately, our proof is by contradiction, and therefore non-constructive. The fact that we cannot make our estimates explicit, even in principle, comes from this fact.

Proof. Suppose not. Then, there is $\varepsilon_0 > 0$ and sequences points $\{x_a^{(n)}\}_{a=1}^{N_n}$ with $N_n \rightarrow \infty$ and

$$\sum_{a \neq b} \frac{1}{|x_b^{(n)} - x_a^{(n)}|} < \frac{(1 - \varepsilon_0)N_n}{|x_b^{(n)}|} \quad (3.1)$$

for all n and all $1 \leq b \leq N_n$.

Equation (3.1) is invariant under rotations and scaling of the x 's as well as relabelling. Thus, without loss we can suppose that

$$x_1^{(n)} = (1, 0, 0) = x_0, \\ |x_a^{(n)}| \leq 1 \quad \text{all } 1 \leq a \leq N_n.$$

The measures

$$\rho^{(n)} = N_n^{-1} \sum_{a=1}^{N_n} \delta(x - x_a^{(n)})$$

are probability measures on the unit ball. Thus, by passing to a subsequence if necessary, we can suppose that ρ_n converges in the $C(\mathbb{R}^3)$ -weak topology to a probability measure $d\rho$. We will show that $d\rho$ violates Theorem 2.1. If y is a limit point of $x_i^{(n)}$ and $g(z) = (|z|^2 + M^2)^{-1/2}$, then since g is C^1 with bounded derivatives

$$\lim \int g(x - x_i^{(n)}) d\rho_n(x) = \int g(x - y) d\rho(x).$$

Thus, by (3.1):

$$\int g(x - y) d\rho(x) \leq (1 - \varepsilon_0) |y|^{-1}.$$

By the monotone convergence theorem, we can take M to zero to obtain

$$\int |x - y|^{-1} d\rho(x) \leq (1 - \varepsilon) |y|^{-1}. \quad (3.2)$$

We have just proven (3.2) for any y in the limit set of the $\{x_a^{(n)}\}$. Any $y \in \text{supp } \rho$ is such a limit point so (3.2) holds for all $y \in \text{supp } \rho$. Thus we will have a contradiction with Theorem 2.1 if we show that $\rho \neq \delta_0$. But since $x_1^{(n)} = x_0$, we have (3.2) for $y = x_0$, i.e.,

$$\int |x - x_0|^{-1} d\rho(x) \leq (1 - \varepsilon). \quad (3.3)$$

Since, for $d\rho = \delta_0$, the left side is 1, we can conclude that $d\rho \neq \delta_0$. $\square$

We will actually need an extension of Theorem 3.1 to potentials cut off at short distances, but in a way that may seem unnatural at first. Define

$$G_\alpha(x, y) = \begin{cases} |x - y|^{-1} & \text{if } |x - y| \geq \alpha |x| \\ \alpha^{-1} |x|^{-1} & \text{if } |x - y| \leq \alpha |x|. \end{cases} \quad (3.4)$$

For a set of points $\{x_a\}_{a=1}^N$ in $\mathbb{R}^3$, define $|x|_\infty \equiv \sup_a |x_a|$.

Theorem 3.2. *Let $\alpha_N \rightarrow 0$ as $N \rightarrow \infty$. Then, for any ε , there exists N_0 and $\delta > 0$ so that, for any $N \geq N_0$ and any set of points $\{x_a\}_{a=1}^N$, there is a point x_a with $|x_a| \geq \delta |x|_\infty$ and*

$$\sum_{j \neq a} G_{\alpha_N}(x_a, x_j) \geq \frac{(1 - \varepsilon)N}{|x_a|}. \quad (3.5)$$

Proof. G is defined to be invariant under scaling (which is why we took the cutoff to be $\alpha|x|$, not just α) and rotations. Also, the condition $|x_a| \geq \delta |x|_\infty$ has the same invariance. Thus, if the result is false, we can find $\varepsilon_0 > 0$, $\delta_N \rightarrow 0$ and a sequence with $|x|_\infty = 1$; $x_1^{(n)} = (1, 0, 0)$ so that (3.5) fails. Taking the limit, we get the same contradiction as in the proof of Theorem 3.1. $\square$

4. A Partition of Unity

As noted in Sect. 1, the key element in the proof of Sigal, which we will mimic, is the construction of a partition of unity. Here we will construct such a partition which we will use in the next section. The preliminaries in the last section will be relevant precisely in order to be sure that certain sets cover $\mathbb{R}^{3N}$.

Theorem 4.1. *For all $\varepsilon > 0$, there exists N_0 and $\delta > 0$, and for each $N \geq N_0$ and each $R > 0$, a family $\{J_a\}_{a=0}^N$ of C^∞ functions on $\mathbb{R}^{3N}$ so that:*

- (1) J_0 is totally symmetric, $\{J_a\}_{a \neq 0}$ is symmetric in $\{x_b\}_{b \neq a}$.
 - (2) $\sum_a J_a^2 = 1$.
 - (3) $\text{supp } J_0 \subset \{\{x_a\} \mid |x|_\infty < R\}$.
 - (4) $\text{supp } J_a \subset \{x \mid |x|_\infty \geq (1 - \varepsilon)R, |x_a| \geq (1 - 2\varepsilon)\delta|x|_\infty \text{ and } \sum_{b \neq a} |x_b - x_a|^{-1} \geq (1 - 2\varepsilon)N|x_a|^{-1}\}$.
 - (5) For a constant C , depending only on ε , $\sum_a |\nabla J_a|^2 \leq CN^{1/2}(\ln N)^2|x|_\infty^{-1}R^{-1}$.
- (4.1)

Proof. Without loss, we can take $R = 1$ since the result for $R = 1$ implies the result for all R by scaling. Moreover, we can prove (4.1) with $|x|_\infty^{-1}$ replaced by $|x|_\infty^{-2}$. For the left-hand side of (4.1) (when $R = 1$) is supported in the region where $|x|_\infty \geq (1 - \varepsilon)$, and in that region $|x|_\infty^{-2} \leq (1 - \varepsilon)^{-1}|x|_\infty^{-1}$.

Next, we note that instead of finding J_a 's obeying (1)–(5), it suffices to find F_a 's obeying (1'), (2'), (3), (4) and (5'):

$$(2') \sum_a F_a^2 \geq \frac{1}{2},$$

$$(5') \left(\sum_a (\nabla F_a)^2 \right) / \left(\sum_a F_a^2 \right) \leq CN^{1/2}(\ln N)^2|x|_\infty^{-2},$$

and for any permutation π :

$$(1') F_{\pi(a)}(x_{\pi(1)}, \dots, x_{\pi(N)}) = F_a(x_1, \dots, x_N).$$

For if $J_a = F_a / \left(\sum_a F_a^2 \right)^{1/2}$, then J_a has the same symmetry and support properties as F has, and

$$\sum_a (\nabla J_a)^2 \leq \sum_a (\nabla F_a)^2 / \sum_a F_a^2. \quad (4.2)$$

To understand (4.2), think of $F = (F_0, \dots, F_N)$ as a function from $\mathbb{R}^{3N}$ to $\mathbb{R}^{N+1}$, in which case J is the “angular” part of F , and (4.2) is a standard inequality on the gradient of the angular part.

Now we concentrate on constructing the F_a 's. Let ψ be a C^∞ function on $[0, \infty)$ with

$$\begin{aligned} \psi(y) &= 0 & y &\leq 1 - 2\varepsilon \\ &= 1 & y &\geq 1 - \varepsilon \\ &\in [0, 1] & \text{all } y \end{aligned}$$

and define $\varphi = \psi^2$. Let $\alpha_N \equiv (\ln N)^{-1}$ and choose N_0, δ as given by Theorem 3.2. As a preliminary, take F_0, F_a as follows. These functions are not C^∞ , but are continuous and are C^1 off the set $\{x \mid |x_a| = |x_b| \text{ for some } a \neq b\} \cup \{x \mid |x_a - x_b| = \alpha_N|x_a| \text{ some } a, b\}$ with discontinuities of the gradients allowed on that set:

$$F_0(x) = 1 - \psi(|x|_\infty),$$

$$F_a(x) = \varphi(|x|_\infty) \varphi(|x_a|/\delta|x_\infty|) \varphi\left(N^{-1}|x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b)\right),$$

where G_α is given by (3.3). The symmetry condition (1) is obvious, and (3) holds since $\varphi(y) = 1$ if $y \geq 1$.

$F_a \neq 0$ implies that $|x|_\infty \geq 1 - 2\varepsilon$, $|x_a| \geq (1 - \varepsilon)\delta|x|_\infty$, and

$$\sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \geq (1 - 2\varepsilon)N|x_a|^{-1},$$

since $\varphi(y) \neq 0$ implies $|y| \geq 1 - 2\varepsilon$. Since $|x - y|^{-1} \geq G_\alpha(x, y)$, we have proven (4). That leaves the key conditions (5') and (2').

For (2'), we use Theorem 3.2. This guarantees us that there is an a where the last two factors in F_a are 1, and so there is an a with $F_a(x) = \varphi(|x|_\infty)$. Since

$$\theta^2 + (1 - \theta)^2 \geq \frac{1}{2}$$

for all θ , for this a , $F_0(x)^2 + F_a(x)^2 \geq \frac{1}{2}$, proving (2').

As a preliminary to (5'), we want to note that, for some constant C and all γ :

$$|\varphi'|^2 \leq \gamma + C\gamma^{-1}|\varphi|^2. \quad (4.3)$$

For

$$|\varphi'|^2 = 4|\psi|^2|\psi'|^2 \leq \gamma + 4\gamma^{-1}|\psi|^4|\psi'|^4,$$

proving (4.3) with $C = 4\|\psi'\|_\infty^4$.

Away from points where $|x_a| = |x_b|$ for some $a \neq b$,

$$\nabla\varphi(|x|_\infty) = \varphi'(|x|_\infty)\nabla|x|_\infty.$$

Since $|x|_\infty$ is some $|x_a|$ and

$$\nabla_a|x_b| = \delta_{ab}x_a/|x_a|,$$

we see that

$$|\nabla\varphi(|x|_\infty)| = \varphi'(|x|_\infty). \quad (4.4)$$

Similarly,

$$\nabla\varphi(|x_a|/\delta|x|_\infty) = \varphi'(|x_a|/\delta|x|_\infty) \left[-\frac{|x_a|}{\delta|x|_\infty^2} \nabla|x|_\infty + \delta^{-1}|x|_\infty^{-1} \nabla|x_a| \right]. \quad (4.5)$$

Finally, if

$$\eta_a = \varphi(N^{-1}|x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b)),$$

then, for $b \neq a$

$$|\nabla_b \eta_a| \leq \varphi'(N^{-1}|x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b)) N^{-1}|x_a| |G_{\alpha_N}(x_a, x_b)|^2,$$

since

$$\nabla_b G_\alpha = [G_\alpha^2](x_b - x_a)/|x_b - x_a| \quad \text{or} \quad 0.$$

Recall that $G_\alpha(x, y) \leq \alpha^{-1}|x|^{-1}$, so since $\alpha_N^{-1} = \ln N$,

$$\sum_{b \neq a} |\nabla_b \eta_a|^2 \leq \left[\varphi' \left(N^{-1}|x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \right) \right]^2 N^{-2}|x_a|^2 (\ln N)^3 |x_a|^{-3} \sum_{b \neq a} G_{\alpha_N}(x_a, x_b).$$

However, $\text{supp } \varphi' \subset [1 - 2\varepsilon, 1 - \varepsilon]$, so on $\text{supp } \varphi' \left(N^{-1}|x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \right)$, we

have that

$$\sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \leq |x_a|^{-1} N.$$

Thus

$$\sum_{b \neq a} |\nabla_b \eta_a|^2 \leq N^{-1} (\ln N)^3 |x_a|^{-2} \left[\varphi' \left(N^{-1} |x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \right) \right]^2. \quad (4.6)$$

As a final gradient estimate,

$$\begin{aligned} |\nabla_a \eta_a| &\leq \varphi' \left(N^{-1} |x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \right) \left[N^{-1} \sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \right. \\ &\quad \left. + N^{-1} \sum_{b \neq a} |x_a| G_{\alpha_N}(x_a, x_b)^2 \right], \end{aligned}$$

since $|\nabla_a |x_a|| = 1$ and

$$|\nabla_a G_\alpha(x_a, x_b)| = G_\alpha(x_a, x_b)^2 \quad \text{or} \quad \alpha G_\alpha(x_a, x_b)^2$$

and $\alpha \leq 1$. Thus, since $|x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \leq N$ when $\varphi'(\cdot) \neq 0$,

$$|\nabla_a \eta_a| \leq \varphi' \left(N^{-1} |x_a| \sum_{b \neq a} G_{\alpha_N}(x_a, x_b) \right) (1 + (\ln N)) |x_a|^{-1}. \quad (4.7)$$

Since $|x_a| \geq (1 - 2\varepsilon) \delta |x|_\infty$ on $\text{supp } F_a$, we see that, for $a \neq 0$.

$$\sum_b |\nabla_b F_a|^2 \leq c_1 (\gamma + c_2 \gamma^{-1} F_a^2) (\ln N)^2 |x|_\infty^{-2},$$

by using (4.3)–(4.7). Thus

$$\sum_{a,b} |\nabla_b F_a|^2 \leq c_1 \left(\gamma N + c_2 \gamma^{-1} \sum_a F_a^2 \right) (\ln N)^2 |x|_\infty^{-2} + c_3 |x|_\infty^{-2}.$$

Since $\sum F_a^2 \geq \frac{1}{2}$, we can take $\gamma = N^{-1/2}$ and obtain:

$$\sum_{a,b} |\nabla_b F_a|^2 \leq c_4 N^{1/2} (\ln N)^2 |x|_\infty^{-2} \sum_a F_a^2,$$

as required.

The J 's constructed in this way are continuous but are only piecewise C^1 . By convoluting with a smooth, totally symmetric function of very small support, we can arrange for C^∞ J 's which still obey the required properties. $\square$

Remark. By using $\varphi = \psi^m$ in the above construction for m suitable, we can reduce $N^{1/2}$ to any desired positive power of N .

5. The Main Theorem

Here we will prove Theorem 1.1. Given the construction in the last section, this follows Sigal [12] fairly closely. Pick $\varepsilon > 0$. We shall prove $\lim N(Z)/Z \leq (1 - 3\varepsilon)^{-1}$.

Let $\{J_a\}$ be as in Theorem 4.1, and let

$$L = \sum_{a=0}^N |\nabla J_a|^2.$$

By the IMS localization formula (see Chap. 3 of [3]),

$$H = \sum_{a=0}^N J_a H J_a - L = \sum_{a=0}^N J_a (H - L) J_a. \quad (5.1)$$

By condition (5) of Theorem 4.1,

$$L \leq CN^{1/2} (\ln N)^2 |x|_\infty^{-1} R^{-1}.$$

For $a > 0$, $\text{supp } J_a \subset \{x \mid |x_a| \geq (1 - 2\varepsilon)\delta |x|_\infty\}$, and thus:

$$J_a L J_a \leq c_1 N^{1/2} (\ln N)^2 |x_a|^{-1} R^{-1} \quad (C_1 = c\delta^{-1}(1 - 2\varepsilon)^{-1}). \quad (5.2)$$

Since $\cup \text{supp}(\nabla J_a) \subset \{x \mid |x|_\infty \geq (1 - 2\varepsilon)R\}$:

$$J_0 L J_0 \leq c_2 N^{1/2} (\ln N)^2 R^{-2} \quad (c_2 = (1 - 2\varepsilon)^{-1} C) \quad (5.3)$$

Let $H_a(N - 1, Z)$ be the $(N - 1)$ electron Hamiltonian obtained by removing from $H(N, Z)$ all terms involving x_a , so:

$$H(N, Z) = H_a(N - 1, Z) - \Delta_a - |x_a|^{-1} Z + \sum_{b \neq a} |x_b - x_a|^{-1}.$$

Since $H_a(N - 1, Z) \geq E(N - 1, Z)$ and $-\Delta_a \geq 0$, taking into account (5.2) and the support property of J_a , we have that

$$J_a (H(N, Z) - L) J_a \geq J_a [E(N - 1, Z) + |x_a|^{-1} d(Z, N, R)] J_a, \quad (5.4a)$$

where

$$d(Z, N, R) = -Z - c_1 N^{1/2} (\ln N)^2 R^{-1} + (1 - 2\varepsilon)N. \quad (5.4b)$$

R_N has not been introduced up to now.

By solving for a Bohr atom (and this is where the Pauli principle enters):

$$\sum_{i=1}^N (-\Delta_i - Z|x_i|^{-1}) \geq -c_3 Z^2 N^{1/3},$$

so since $|x_a - x_b| \leq 2R$ on $\text{supp } J_0$:

$$J_0 (H(N, Z) - L) J_0 \geq J_0 [-c_3 Z^2 N^{1/3} - c_2 N^{1/2} (\ln N)^2 R^{-2} + \frac{1}{4} R^{-1} N(N - 1)] J_0. \quad (5.5)$$

Choose $R = N^{-2/5}$. Then, for $N \geq (1 - 3\varepsilon)^{-1} Z$ and large Z , $d(Z, N, R) > 0$ since $\frac{1}{2} + \frac{2}{5} < 1$. Moreover, $J_0 (H(N, Z) - L) J_0 \geq 0 \geq J_0 E(N - 1, Z) J_0$ since $N^{12/5}$ dominates $N^{7/3}$ and $N^{13/10} (\ln N)^2$ and $E(N - 1, Z) \leq 0$.

Thus,

$$H(H, Z) \geq E(N - 1, Z)$$

if $N \geq (1 - 3\varepsilon)^{-1} Z$ and Z is large, i.e., for Z large

$$N(Z) \leq (1 - 3\varepsilon)^{-1} Z.$$

Since ε is arbitrary:

$$\overline{\lim} N(Z)/Z \leq 1.$$

It is well known (see [15,13]) that $H(Z, Z)$ has bound states, i.e., that $N(Z) \geq Z$. $\square$

Remark. Without the Pauli principle, $Z^2 N^{1/3}$ becomes $Z^2 N$, so one must take $R_N = cN^{-1}$, in which case the localization term $N^{1/2}(\ln N)^2 R_N^{-1}$ in (5.4b) becomes uncontrollable. Our proof must, of course, fail without the Pauli principle because of the result in [1].

6. Extensions

Our result extends easily to accommodate arbitrary magnetic fields (the same for all electrons) and/or a finite nuclear mass.

The exact form of the electron kinetic energy entered only in two places: in the IMS localization formula and in the positivity of $-\Delta$, both of which hold in an arbitrary magnetic field. We also used the Bohr atom binding energy, but that only decreases in a magnetic field (i.e., $-c_3 N^2 Z^{1/3}$ is a lower bound for all fields). Thus, we obtain a magnetic field independent bound $\tilde{N}(Z)$ with

$$\tilde{N}(Z)/Z \rightarrow 1 \quad \text{as } Z \rightarrow \infty.$$

As for finite nuclear mass, let x_0 be the nuclear coordinate, and use $J_a(x_b - x_0)$ in place of $J_a(x_b)$. With this change, the nuclear coordinates pass through all proofs with essentially no change at all.

Acknowledgement. This work was begun while I.S. was at the Weizmann Institute, and B.S. would like to thank H. Dym and I. Sigal for the hospitality of that Institute. W.T. would like to thank E. Lieb for the hospitality of Princeton University, and M. Goldberger and R. Vogt for the hospitality of Caltech. An announcement appeared in [8].

References

1. Benguria, R., Lieb, E.: Proof of the stability of highly negative ions in the absence of the Pauli principle. *Phys. Rev. Lett.* **50**, 1771 (1983)
2. Choquet, G.: Sur la fondements de la théorie finie du potentiel. *C.R. Acad. Sci. Paris* **244**, 1606 (1957)
3. Cycon, H., Froese, R., Kirsch, W., Simon, B.: *Schrödinger operators with application to quantum mechanics and global geometry*. Berlin, Heidelberg, New York: Springer 1987
4. Evans, G.: On potentials of positive mass, I. *Trans. AMS* **37**, 226 (1935)
5. Helms, L.: *Introduction to potential theory*. New York: Wiley 1966
6. Lieb, E.: Atomic and molecular ionization. *Phys. Rev. Lett.* **52**, 315 (1984)
7. Lieb, E.: Bound on the maximum negative ionization of atoms and molecules. *Phys. Rev.* **A29**, 3018–3028 (1984)
8. Lieb, E., Sigal, I. M., Simon, B., Thirring, W.: Asymptotic neutrality of large-Z ions. *Phys. Rev. Lett.* **52**, 994 (1984)
9. Ruskai, M.: Absence of discrete spectrum in highly negative ions. *Commun. Math. Phys.* **82**, 457–469 (1982)
10. Ruskai, M.: Absence of discrete spectrum in highly negative ions, II. *Commun. Math. Phys.* **85**, 325–327 (1982)

11. Sigal, I. M.: Geometric methods in the quantum many-body problem. Nonexistence of very negative ions. *Commun. Math. Phys.* **85**, 309–324 (1982)
12. Sigal, I. M.: How many electrons can a nucleus bind? *Ann. Phys.* **157**, 307–320 (1984)
13. Simon, B.: On the infinitude or finiteness of the number of bound states of an N -body quantum system, I. *Helv. Phys. Acta* **43**, 607–630 (1970)
14. Vasilescu, F.: Sur la contribution du potentiel à travers des masses et la démonstration d'une lemme de Kellogg. *C.R. Acad. Sci. Paris* **200**, 1173 (1935)
15. Zhislin, G.: Discussion of the spectrum of Schrödinger operator for systems of many particles. *Tr. Mosk. Mat. Obs.* **9**, 81–128 (1960)
16. Baumgartner, B.: On Thomas–Fermi–von Weizsäcker and Hartree energies as functions of the degree of ionisation. *J. Phys.* **A17**, 1593–1602 (1984)
17. Baxter, J.: Inequalities for potentials of particle systems, III. *J. Math.* **24**, 645–652 (1980)

Communicated by A. Jaffe

Received December 7, 1987

Universal nature of van der Waals forces for Coulomb systems

Elliott H. Lieb

*Department of Mathematics and Department of Physics, Princeton University,
Jadwin Hall, P.O. Box 708, Princeton, New Jersey 08544*

Walter E. Thirring

*Institut für Theoretische Physik, Universität Wien, Boltzmannngasse 5, A-1090 Vienna, Austria
(Received 2 December 1985)*

The nonrelativistic Schrödinger equation is supposed to yield a pairwise R^{-6} attractive interaction among atoms or molecules for large separation, R . Up to now this attraction has been investigated only in perturbation theory or else by invoking various assumptions and approximations. We show rigorously that the attraction is at least as strong as R^{-6} for any shapes of the molecules, independent of other features such as statistics or sign of charge of the particles. More precisely, we prove that two neutral molecules can always be oriented such that the ground-state energy of the combined system is less than the sum of the ground-state energies of the isolated molecules by a term $-cR^{-6}$ provided R is larger than the sum of the diameters of the molecules. When several molecules are present, a pairwise bound of this kind is derived. In short, we prove that in the quantum mechanics of Coulomb systems everything binds to everything else if the nuclear motion is neglected.

I. INTRODUCTION

Our purpose is to explore and to answer some elementary but fundamental questions about the binding of neutral atoms and molecules. To simplify matters, we shall use the infinite nuclear-mass approximation in which the nuclei are held fixed, but many of our mathematical constructions can, with additional work and appropriate changes, be carried over to the more realistic case of dynamic nuclei. In our fixed-nuclei approximation we do *not* assume that the nuclei are necessarily in the configuration that minimizes the energy of the molecule.

Consider two neutral molecules (or atoms) labeled α and β , with respective diameters $2r^\alpha$ and $2r^\beta$, and whose centers are separated by a distance $R^{\alpha\beta} > r^\alpha + r^\beta$. (The precise definition of r^α , r^β , and $R^{\alpha\beta}$ will be given in Sec. III.) Let the ground-state energies of the isolated molecules and of the combined system be e^α , e^β , and $e(R^{\alpha\beta})$, respectively. The question we shall address is this: Is it possible to orient the nuclear coordinates at the two neutral molecules with respect to each other (with $R^{\alpha\beta}$ fixed) so that after an appropriate readjustment of the electronic wave function

$$e(R^{\alpha\beta}) < e^\alpha + e^\beta ? \quad (1.1)$$

In particular, is there an upper bound of the van der Waals form

$$e(R^{\alpha\beta}) \leq e^\alpha + e^\beta - C(R^{\alpha\beta})^{-6} \quad (1.2)$$

for a constant $C > 0$ which depends on the intrinsic properties of the two molecules, but not on $R^{\alpha\beta}$?

We shall prove, using a variational argument, that Eq. (1.2) is true; in other words, we prove that "everything binds to everything else" when the nuclear kinetic energy is neglected. First, several remarks are in order.

(1) Equation (1.2), or even Eq. (1.1), implies binding in the fixed-nuclei approximation. When the nuclear kinetic energy is added, the uncertainty principle may destroy the binding, as is probably the case for He_2 . Thus we can only say that sufficiently heavy isotopes will always bind.

(2) Density-functional theories (at least the ones known to us) fail to predict Eq. (1.2). Although a density functional that predicts Eq. (1.2) exists in principle,^{1,2} no one has actually constructed one. In Thomas-Fermi theory even Eq. (1.1) fails because Teller's theorem³ states that in Thomas-Fermi theory $e(R^{\alpha\beta}) > e^\alpha + e^\beta$ always. When gradient corrections are added, as in Thomas-Fermi-von Weizsäcker theory, Eq. (1.1) holds³ when $R^{\alpha\beta} \approx r^\alpha + r^\beta$ but Eq. (1.2) fails when $R^{\alpha\beta} \gg r^\alpha + r^\beta$. The reason for this failure of (local) density-functional theory (as explained in Refs. 2 and 3) is the following.

The R^{-6} attraction comes from a dipole-dipole interaction but (in the combined system) there is almost no static dipole moment in each molecule (if both molecules were free of static dipole moments in their ground states). The interaction energy of dipole moments d^α and d^β on the respective molecules is proportional to $-d^\alpha d^\beta (R^{\alpha\beta})^{-3}$. Density-functional theory, since it deals only with single-particle densities, can produce these only as static moments and at an energy cost of $\frac{1}{2}c^\alpha (d^\alpha)^2 + \frac{1}{2}c^\beta (d^\beta)^2$. Thus, when $R^{\alpha\beta} > (c^\alpha c^\beta)^{1/6}$ the optimum choice is $d^\alpha = d^\beta = 0$ and there is no attraction.

The true source of the R^{-6} term in Eq. (1.2) is a correlation effect between the electrons in molecule α and those in molecule β . It is essential to think of electrons as particles and not as a simple fluid. In the language of quantum mechanics the molecules make a virtual transition (simultaneously and not separately) to an excited state. The energy to create d^α and d^β is then (from second-order perturbation theory) proportional to $(d^\alpha d^\beta)^2$. The minimum with respect to d^α and d^β for fixed $R^{\alpha\beta}$ of

$(d^\alpha d^\beta)^2 - d^\alpha d^\beta (R^{\alpha\beta})^{-3}$ is the required $-(R^{\alpha\beta})^{-6}$. Apart from the fact that quantum mechanics is needed in order to give meaning to the concept of the ground state, the effect is classical insofar as no interference effect is involved. There is simply a coherence in the motion of the electrons in each molecule so that the (time) average dipole-dipole interaction is not zero even though the average dipole moment of each molecule is essentially zero.

(3) When one studies the question of the additivity of the van der Waals forces between pairs of molecules the correlation effect appears even more strikingly. The Coulomb interaction correlates the dipole moments $\mathbf{d}^\alpha$ and $\mathbf{d}^\beta$ with the displacement $\mathbf{R}^{\alpha\beta}$ between them in such a way as to give the directional factor $-3(\mathbf{d}^\alpha \cdot \mathbf{R}^{\alpha\beta})(\mathbf{d}^\beta \cdot \mathbf{R}^{\alpha\beta}) + (\mathbf{d}^\alpha \cdot \mathbf{d}^\beta) |\mathbf{R}^{\alpha\beta}|^2$ as function of direction its minimal value, $-2|\mathbf{d}^\alpha| |\mathbf{d}^\beta| |\mathbf{R}^{\alpha\beta}|^2$, simultaneously for all pairs. Thus the nonstatic dipole moments $\mathbf{d}^\alpha$ do not depend on the *single* molecules, but only on the *pairs* of molecules which interact with each other. As a consequence, our bounds on the effective interaction potentials will add together like scalar potentials and not like dipole potentials. This does not, of course, imply that the true effective interaction potential has this property.

(4) The analysis in remark (2) was based on the Coulomb potential, but presumably this is not essential. If the electron-electron interaction were r^{-p} instead of r^{-1} , we should expect the appropriate modification of Eq. (1.2) would have R^{-2p-4} in place of R^{-6} . We shall not pursue this aspect of the problem, however, and will confine our attention to the Coulomb potential.

(5) It is not at all essential that the dynamic particles are electrons. They could be any mixture of bosons and fermions. Also, for example, matter and antimatter will bind in the infinite nuclear-mass approximation—which carries the physical implication that there is no quantum-mechanical Coulomb barrier to the annihilation process. There are several variational calculations of the dependence of the energy on the nuclear separation, but with different conclusions.⁴⁻⁶

(6) No assumption is made about the spherical symmetry of the two molecules and they could have permanent electric dipole or higher-pole moments in their ground states (but not monopole moments). Parity conservation does not preclude this since the nuclear coordinates are fixed. A feature of Eq. (1.2) is that it is independent of any assumption about the permanent moments; if any exist then the binding could be stronger, but not weaker than R^{-6} .

(7) There is, of course, an enormous amount of literature about van der Waals forces (see, e.g., Refs. 7-9). In a certain sense our results are thus not new, but from another point of view they are new. The drawback to the usual theories is that they are always based on perturbation theory in two ways: (i) One assumes that R is sufficiently large so that the $1/r$ Coulomb interaction can be expanded up to the dipole-dipole order, and all higher terms ignored. Although it can be shown that this expansion is asymptotically exact,¹⁰ we are usually not told how large this R has to be. (ii) One uses second-order quantum perturbation theory—and this is usually calculated

with some unverified assumptions about the excited-state molecular or atomic wave functions. Our point is that none of this is necessary. While we make no pretense to getting the correct constant C in Eq. (1.2), we do get a lower bound to the binding energy of the correct form (when $R > r^\alpha + r^\beta$) by a fairly simple and direct variational argument.

(8) Another point about the standard theory that needs to be addressed is the well-known effect of retardation discovered by Casimir and Polder¹¹ and elaborated by Lifshitz.¹² The R^{-6} term in Eq. (1.2) is replaced by R^{-7} when R is "large." However, large means (Bohr radius)/(fine-structure constant) and this is huge compared to $r^\alpha + r^\beta$ (for molecules that are not too large). Thus, for distances of major interest for binding, it is physically correct to use the Schrödinger equation without retardation, and hence Eq. (1.2) is meaningful. For small R , nuclear recoil effects may play a role (see Ref. 13).

The calculation and notation in this paper will seem to the reader to be complex. Actually, the complexity is more apparent than real, and a few words about the strategy of our proof may be helpful. The implementation of the strategy will be given in detail later.

We start with several small units (molecules) which we call clusters. There are $\mathcal{C}$ clusters and $\alpha=1,2,\dots,\mathcal{C}$ designates the cluster. The ground-state energy and wave function of each cluster is e^α and ϕ_0^α . A conveniently chosen point in each cluster, called the "center," is denoted by $\mathbf{R}^\alpha$. The problem is to construct a variational trial function ψ (of all the variables in the system) whose energy is lower than $\sum e^\alpha$ by an amount $\text{const} \times \sum (R^{\alpha\beta})^{-6}$.

Step 1: Apply a cutoff to each ϕ_0^α in a large ball centered at $\mathbf{R}^\alpha$, in such a way that the balls are disjoint. By the variational principle the energy must increase, but since ϕ_0^α decays exponentially, this increase of the energy will only be by an amount proportional to $\sum \exp(-R_\alpha)$, where R_α is the radius of the cutoff. The cutoff function is denoted by ϕ^α .

Step 2: Let $\psi_m^\alpha = \mathbf{m} \cdot \sum_{i=1}^{n^\alpha} \nabla_i \phi^\alpha$ where n^α is the number of electrons in cluster α and $|\mathbf{m}|=1$. The trial function is given by

$$\psi = \prod_{\alpha=1}^{\mathcal{C}} \phi^\alpha + \sum_{\substack{\alpha,\beta \\ (1 \leq \alpha < \beta \leq \mathcal{C})}} \lambda_{\alpha\beta} \psi_{\mathbf{m}_{\alpha\beta}}^\alpha \psi_{\mathbf{n}_{\alpha\beta}}^\beta \prod_{\gamma \neq (\alpha,\beta)}'' \phi^\gamma. \quad (1.3)$$

The $\lambda_{\alpha\beta}$ are adjustable constants and the $\mathbf{m}_{\alpha\beta}$ and $\mathbf{n}_{\alpha\beta}$ are adjustable vectors. Since $\langle \psi_{\mathbf{m}_{\alpha\beta}}^\alpha | \phi^\alpha \rangle = 0$, the normalization $\langle \psi | \psi \rangle$ will be of the form

$$1 + \sum_{\substack{\alpha,\beta \\ (\alpha < \beta)}} \lambda_{\alpha\beta}^2 \times \text{const}.$$

However, $\langle \psi | H \psi \rangle$ will have a cross term proportional to the $\lambda_{\alpha\beta}$ coming from the intercluster Coulomb potential. We should like $\langle \psi | H \psi \rangle$ to be of the form

$$\left[\sum_\alpha e^\alpha \right] \langle \psi | \psi \rangle + \sum_{\alpha,\beta} \lambda_{\alpha\beta}^2 \times \text{const} \\ + \sum_{\alpha,\beta} \lambda_{\alpha\beta} (R^{\alpha\beta})^{-3} \times \text{const}.$$

Minimizing the ratio $\langle \psi | H \psi \rangle / \langle \psi | \psi \rangle$ with respect to

each $\lambda_{\alpha\beta}$ would then lead (more or less) to $\lambda_{\alpha\beta} \sim (R^{\alpha\beta})^{-3}$ and hence to

$$E < \sum_{\alpha} e^{\alpha} - \text{const} \times \sum_{\substack{\alpha, \beta \\ (\alpha < \beta)}} (R^{\alpha\beta})^{-6}.$$

Step 3: Unfortunately, the intercluster Coulomb interaction is not simply $\lambda_{\alpha\beta}(R^{\alpha\beta})^{-3}$. Its calculation is complicated by the fact that each cluster is not spherically symmetric. To avoid this difficulty and to get a rigorous bound to the $\lambda_{\alpha\beta}$ term, we consider all spatial orientations of the clusters and orientations of the $\mathbf{m}_{\alpha\beta}$ and $\mathbf{n}_{\alpha\beta}$ (which enters in the definition of ψ). By averaging over the orientations of the clusters and also averaging over certain selected orientations of the $\mathbf{m}$'s and $\mathbf{n}$'s so that $\langle \psi | H \psi \rangle \leq (\text{average over orientations}) \leq (\sum e^{\alpha} + \text{van der Waals term}) \langle \psi | \psi \rangle$.

In any particular case, the wave function ψ which we construct is not the best possible one within the framework of functions of the type (1.3). We use it because we are striving for simplicity and generality—not for good constants in the energy bound.

II. THE GENERAL PROCEDURE

First, some more notation is needed. There are $\mathcal{C}$ clusters indexed by α , $\alpha = 1, 2, \dots, \mathcal{C}$ and $\mathbf{R}^{\alpha} \in \mathbb{R}^3$ denotes the center of the cluster α . $\mathbf{X}_i^{\alpha}$, $i = 1, 2, \dots, M^{\alpha}$ are the coordinates of the nuclei in α relative to $\mathbf{R}^{\alpha}$. This is also denoted collectively by $\mathbf{X}^{\alpha}$. Similarly, $\mathbf{Z}_i^{\alpha} \in \mathbb{N}$ and $\mathbf{Z}^{\alpha}$ denote the charges of these nuclei. We suppose that all the nuclear coordinates $\mathbf{X}^{\alpha}$ are contained in a ball, B^{α} , of radius $R_{\alpha} > 0$, centered at $\mathbf{R}^{\alpha}$ and that the $\mathcal{C}$ balls are disjoint. In fact, we can define R_{α} by

$$R_{\alpha} = \frac{1}{2} \min_{\substack{\alpha, \beta \\ \beta \neq \alpha}} \{ |\mathbf{R}^{\alpha\beta}| \}, \quad (2.1)$$

with $\mathbf{R}^{\alpha\beta} \equiv \mathbf{R}^{\alpha} - \mathbf{R}^{\beta}$. Later on we shall specify how large R_{α} must be in order that our bounds have a simple form.

If only cluster α were present, it would have $n^{\alpha} = \sum_i \mathbf{Z}_i^{\alpha}$ electrons to be neutral. The coordinates of these n^{α} electrons (relative to $\mathbf{R}^{\alpha}$) are denoted by $\mathbf{x}_i^{\alpha}$ and collectively by $\mathbf{x}^{\alpha}$. The wave functions we shall use will be sums of functions of the form $\Phi = \phi^1(\mathbf{x}^1) \phi^2(\mathbf{x}^2) \cdots \phi^{\mathcal{C}}(\mathbf{x}^{\mathcal{C}})$ with $\phi^{\alpha}(\mathbf{x}^{\alpha}) = 0$ if any $\mathbf{x}_i^{\alpha}$ is outside of B^{α} . Each ϕ^{α} satisfies the Pauli principle but Φ does not. (The electron spins should be included but, since they play no essential role, we omit explicit reference

to the spins for simplicity of notation.) However, if Φ is antisymmetrized no cross terms appear in $\langle \Phi | H \Phi \rangle$ or in $\langle \Phi | \Phi \rangle$ (because the B^{α} are disjoint) which means that we can simply ignore the antisymmetrization of Φ . Therefore it makes sense in the full problem to continue to speak of n^{α} electrons being associated with cluster α .

Because of translation invariance, $\mathbf{R}^{\alpha}$ will not appear in the Hamiltonian H^{α} of cluster α but only in the combination $\mathbf{R}^{\alpha\beta} = \mathbf{R}^{\alpha} - \mathbf{R}^{\beta}$ in the interaction between cluster α and cluster β . In atomic units, $\hbar = e = 2m_e = 1$, and, in the above notation, H^{α} is written as

$$\begin{aligned} H^{\alpha} = & \sum_{i=1}^{n^{\alpha}} -\Delta_{i,\alpha} + \sum_{\substack{i,k \\ (1 \leq i < k \leq n^{\alpha})}} |\mathbf{x}_i^{\alpha} - \mathbf{x}_k^{\alpha}|^{-1} \\ & - \sum_{i=1}^{n^{\alpha}} \sum_{j=1}^{M^{\alpha}} \mathbf{Z}_j^{\alpha} |\mathbf{x}_i^{\alpha} - \mathbf{X}_j^{\alpha}|^{-1} \\ & + \sum_{\substack{i,k \\ (1 \leq i < k \leq M^{\alpha})}} \mathbf{Z}_i^{\alpha} \mathbf{Z}_k^{\alpha} |\mathbf{X}_i^{\alpha} - \mathbf{X}_k^{\alpha}|^{-1}. \end{aligned} \quad (2.2)$$

The interaction between two clusters again involves an electron-electron, an electron-nucleus, and a nucleus-nucleus interaction,

$$\begin{aligned} V^{\alpha\beta} = & \sum_{i,j} |\mathbf{x}_i^{\alpha} - \mathbf{x}_j^{\beta} + \mathbf{R}^{\alpha\beta}|^{-1} - \sum_{i,j} |\mathbf{x}_i^{\alpha} - \mathbf{X}_j^{\beta} + \mathbf{R}^{\alpha\beta}|^{-1} \mathbf{Z}_j^{\beta} \\ & - \sum_{i,j} |\mathbf{x}_j^{\beta} - \mathbf{X}_i^{\alpha} + \mathbf{R}^{\alpha\beta}|^{-1} \mathbf{Z}_i^{\alpha} \\ & + \sum_{i,j} |\mathbf{X}_i^{\alpha} - \mathbf{X}_j^{\beta} + \mathbf{R}^{\alpha\beta}|^{-1} \mathbf{Z}_i^{\alpha} \mathbf{Z}_j^{\beta}. \end{aligned} \quad (2.3)$$

The total Hamiltonian consists of H^0 , the sum of the cluster Hamiltonians, and the interaction V between them

$$H(\mathbf{X}^{\alpha}, \mathbf{R}^{\alpha\beta}) = H^0 + V = \sum_{\alpha=1}^{\mathcal{C}} H^{\alpha} + \sum_{\substack{\alpha, \beta \\ (1 \leq \alpha < \beta \leq \mathcal{C})}} V^{\alpha\beta}. \quad (2.4)$$

To formalize the averaging procedure denote by $\mathcal{R}^{\alpha} \mathbf{X}_k^{\alpha}$ a rotation of the coordinate $\mathbf{X}_k^{\alpha}$ around the center $\mathbf{R}^{\alpha}$ and by $d\mathcal{R}^{\alpha}$ the normalized volume element (Haar measure) of the rotations. We denote $\prod_{\alpha} d\mathcal{R}^{\alpha}$ by $d\mathcal{R}$. What we shall show is that for some trial functions $\psi(\mathbf{X}^{\alpha}, \underline{m})$, which in addition to the nuclear coordinates also depend on some collection of polarization vectors, $\underline{m}$, we have for sufficiently large $|\mathbf{R}^{\alpha\beta}|$,

$$\int d\mathcal{R} \sum_{\underline{m}} \left[\langle \psi(\mathcal{R}^{\alpha} \mathbf{X}^{\alpha}, \underline{m}) | H(\mathcal{R}^{\alpha} \mathbf{X}^{\alpha}, \mathbf{R}^{\alpha\beta}) | \psi(\mathcal{R}^{\alpha} \mathbf{X}^{\alpha}, \underline{m}) \rangle - ||\psi(\mathcal{R}^{\alpha} \mathbf{X}^{\alpha}, \underline{m})||^2 \left[\sum_{\alpha=1}^{\mathcal{C}} e^{\alpha} - \sum_{\substack{\alpha, \beta \\ \alpha > \beta}} v_{\alpha\beta}(|\mathbf{R}^{\alpha\beta}|) \right] \right] \leq 0, \quad (2.5)$$

where $v_{\alpha\beta}(R) = C_1 / (R^6 + C_2)$. We shall use the symbol $\sum_{\underline{m}}$ for an average over the polarization directions, whose precise nature will be given in Sec. VI. Since the average

in (2.5) is negative, we reach the conclusion that there are some orientations such that the clusters attract each other at least as much as the van der Waals energy.

III. THE TRIAL FUNCTION

For our results we require the clusters to be clearly separated. By this we mean that the wave functions should not overlap. This, of course, is not true for the ground-state wave functions which decay only exponentially, but may be achieved by a slight adjustment of the wave function which will not cost too much energy. If ϕ_0^α is the (real-valued) ground-state wave function of H^α ,

$$H^\alpha \phi_0^\alpha = e^\alpha \phi_0^\alpha, \quad (3.1)$$

we shall use a function

$$\phi^\alpha(\underline{x}^\alpha, \underline{X}^\alpha) = \phi_0^\alpha(\underline{x}^\alpha, \underline{X}^\alpha) f(\underline{x}^\alpha), \quad (3.2)$$

where

$$f(\underline{x}^\alpha) = \prod_{k=1}^{n_\alpha} \chi(|\underline{x}_k^\alpha|), \quad (3.3)$$

where $\chi(s)$ is a smooth function which is 1 for $s < R_\alpha - 1$ and 0 for $s > R_\alpha$, and such that $|\chi'(s)|, |\chi''(s)| \leq 4$ for $R_\alpha - 1 < s < R_\alpha$ and zero otherwise. For ϕ^α we have

$$H^\alpha \phi^\alpha = e^\alpha \phi^\alpha - \sum_{k=1}^{n_\alpha} (2\nabla_k^\alpha \phi_0^\alpha \cdot \nabla_k^\alpha f + \phi_0^\alpha \Delta_k^\alpha f), \quad (3.4)$$

and, by some partial integrations,

$$\langle \phi^\alpha | H^\alpha \phi^\alpha \rangle = e^\alpha + \sum_{k=1}^{n_\alpha} \int d\underline{x}^\alpha |\phi_0^\alpha|^2 |\nabla_k^\alpha f|^2, \quad (3.5)$$

where we assumed $\langle \phi^\alpha | \phi^\alpha \rangle = 1$.

The radius R_α was defined in (2.1) and it increases with increasing separation of the clusters. It will turn out that the error in the energy caused by replacing ϕ_0^α by ϕ^α depends only on the single-particle electron density $\rho^\alpha(\underline{x})$, defined in the usual way [see (5.2) but omit the spherical average $\int d\underline{R}$]. It is known that $\rho^\alpha(\underline{x})$ decays exponentially, which means that for some r^α and $c > 0$,

$$\rho^\alpha(\underline{x}) < \exp[c(r^\alpha - |\underline{x}|)] \text{ for } |\underline{x}| > r^\alpha.$$

We take r^α as the *definition* of the *cluster radius*. The additional terms in (3.4) and (3.5), which involve only the region $R_\alpha - 1 < |\underline{x}| < R_\alpha$, will thus be exponentially small (as a function of R_α) when $R_\alpha > r^\alpha + 1$.

Next, we need a wave function that describes a polarized cluster and there are many choices possible. We found the most convenient form to be

$$\psi_m^\alpha = \mathbf{m} \cdot \sum_{k=1}^{n_\alpha} \nabla_k^\alpha \phi^\alpha, \quad \mathbf{m} \in \mathbb{R}^3, |\mathbf{m}| = 1. \quad (3.6)$$

It is easily seen to be orthogonal to ϕ^α ,

$$\langle \psi_m^\alpha | \phi^\alpha \rangle = 0, \quad (3.7)$$

and its norm is the expectation value of the square of the total momentum of all electrons in cluster α in the direction of $\mathbf{m}$,

$$\langle \psi_m^\alpha | \psi_m^\alpha \rangle = \int d\underline{x}^\alpha \left| \mathbf{m} \cdot \sum_{k=1}^{n_\alpha} \nabla_k^\alpha \phi^\alpha \right|^2. \quad (3.8)$$

This quantity still depends on the direction $\mathbf{m}$. However, in our results only the scalar product averaged over rotations of the nuclear coordinates will appear and

$$\int d\underline{R} \langle \psi_m^\alpha | \psi_m^\alpha \rangle \equiv \tau^\alpha$$

is independent of $\mathbf{m}$. At the end of the paper we shall collect simple estimates for the various constants such as τ^α which will enter our result. For now we go on to exhibit the trial function ψ for the total system (the products are always in the sense of tensor products)

$$\begin{aligned} \psi &= \prod_{\alpha=1}^{\mathcal{C}} \phi^\alpha + \sum_{\substack{\alpha, \beta \\ (1 \leq \alpha < \beta \leq \mathcal{C})}} \lambda_{\alpha\beta} \phi^1 \cdots \phi^{\alpha-1} \psi_{\mathbf{m}_{\alpha\beta}}^\alpha \phi^{\alpha+1} \cdots \phi^{\beta-1} \\ &\quad \times \psi_{\mathbf{n}_{\alpha\beta}}^\beta \phi^{\beta+1} \cdots \phi^{\mathcal{C}} \\ &\equiv \Phi + \sum_{\substack{\alpha, \beta \\ (\alpha < \beta)}} \lambda_{\alpha\beta} \psi^{\alpha\beta}. \end{aligned} \quad (3.9)$$

Here $\lambda_{\alpha\beta} \in \mathbb{R}$ and $\mathbf{m}_{\alpha\beta}, \mathbf{n}_{\alpha\beta} \in S^2$ are variational parameters to be chosen later. Notice that the polarization vectors $\mathbf{m}$ and $\mathbf{n}$ of the clusters depend on the *pair* of clusters α and β . Because of the orthogonality (3.7) there are no terms linear in the λ 's in the norm of ψ ,

$$\langle \psi | \psi \rangle = 1 + \sum_{\substack{\alpha, \beta \\ (\alpha < \beta)}} \lambda_{\alpha\beta}^2 |\psi_{\mathbf{m}_{\alpha\beta}}^\alpha|^2 |\psi_{\mathbf{n}_{\alpha\beta}}^\beta|^2. \quad (3.10)$$

IV. THE EXPECTATION VALUE OF H^0

In calculating $\langle \psi | H^0 \psi \rangle$ we first note that the term independent of λ is almost e^α except for the exponentially decreasing contribution in (3.5),

$$\langle \Phi | H^0 \Phi \rangle = e^\alpha + \sum_{k=1}^{n_\alpha} \int d\underline{x}^\alpha |\phi_0^\alpha|^2 |\nabla_k^\alpha f|^2 \equiv e^\alpha + b_1^\alpha, \quad (4.1)$$

where the boundary term b_1^α will be less than $c \exp(-R_\alpha)$. The orthogonality (3.7) makes the contributions which are linear in λ vanish,

$$\begin{aligned} \langle \Phi | H^0 \psi^{\alpha\beta} \rangle &= \langle \phi^\alpha \cdot \phi^\beta | H^0 \psi_{\mathbf{m}_{\alpha\beta}}^\alpha \psi_{\mathbf{n}_{\alpha\beta}}^\beta \rangle \\ &= \langle \phi^\alpha | H^0 \psi_{\mathbf{m}_{\alpha\beta}}^\alpha \rangle \langle \phi^\beta | \psi_{\mathbf{n}_{\alpha\beta}}^\beta \rangle = 0. \end{aligned} \quad (4.2)$$

The terms quadratic in λ require some rearrangement because we want to compare $\langle \psi | H^0 \psi \rangle$ with $||\psi||^2 \sum_\alpha e^\alpha$ and not just with $\sum_\alpha e^\alpha$, which we get from (4.1). To achieve this we use the following identity, where we momentarily abbreviate $\sum_{k=1}^{n_\alpha} \mathbf{m} \cdot \nabla_k^\alpha$ by $P^\alpha = -(P^\alpha)^*$,

$$\begin{aligned} \langle \psi^{\alpha\beta} | H^0 | \psi^{\alpha\beta} \rangle &= - \langle \phi^\alpha | P^\alpha H^0 P^\alpha | \phi^\alpha \rangle |\psi_{\mathbf{n}_{\alpha\beta}}^\beta|^2 \\ &= - \frac{1}{2} \langle \phi^\alpha | [P^\alpha, [H^0, P^\alpha]] + H^0 (P^\alpha)^2 + (P^\alpha)^2 H^0 | \phi^\alpha \rangle \\ &\quad \times |\psi_{\mathbf{n}_{\alpha\beta}}^\beta|^2. \end{aligned} \quad (4.3)$$

Using (3.4), $-H^\alpha (P^\alpha)^2 - (P^\alpha)^2 H^\alpha$ indeed yields $e^\alpha ||\psi^{\alpha\beta}||^2$

and some boundary terms, which we call b_2^α , and which are also exponentially decreasing. As regards the double commutator, we first see from (2.2) that the only term in H^α that is not invariant under a common displacement of all electrons is the electron-nucleus attraction

$$[P^\alpha, H^\alpha] = - \sum_{k,j} \mathbf{m}_{\alpha\beta} \cdot \nabla_k^\alpha Z_j^\alpha | \mathbf{x}_k^\alpha - \mathbf{X}_j^\alpha |^{-1}, \quad (4.4)$$

$$[P^\alpha, [P^\alpha, H^\alpha]] = - \sum_{k,j} (\mathbf{m}_{\alpha\beta} \cdot \nabla_k^\alpha)^2 Z_j^\alpha | \mathbf{x}_k^\alpha - \mathbf{X}_j^\alpha |^{-1}. \quad (4.5)$$

A further simplification of the expectation value of this expression results if we carry out the integrations $\int d\mathcal{R}^\alpha \int d\mathcal{R}^\beta$. The integration $\int d\mathcal{R}^\beta$ makes $||\psi_{\alpha\beta}^\beta||^2$ independent of $\mathbf{n}_{\alpha\beta}$ and $\int d\mathcal{R}^\alpha$ makes the expectation value of (4.5) independent of $\mathbf{m}_{\alpha\beta}$. The reason is that after rotating the nuclei there is no distinguished direction left. Since the result is independent of $\mathbf{m}_{\alpha\beta}$, we may freely average over three orthogonal directions of $\mathbf{m}_{\alpha\beta}$ and thus replace $-(\mathbf{m} \cdot \nabla_k^\alpha)^2$ by $-\frac{1}{3} \Delta_k^\alpha$ which just gives $(4\pi/3) \delta(\mathbf{x}_k^\alpha - \mathbf{X}_j^\alpha)$ in (4.5). Upon collecting the contributions we end up with

$$\begin{aligned} \int d\mathcal{R}^\alpha \int d\mathcal{R}^\beta \langle \psi^{\alpha\beta} | H^\alpha | \psi^{\alpha\beta} \rangle &= e^\alpha ||\psi^{\alpha\beta}||^2 + \left[b_2^\alpha + \frac{2\pi}{3} \sum_{k,j} \langle \phi^\alpha | Z_j^\alpha \delta(\mathbf{x}_k^\alpha - \mathbf{X}_j^\alpha) | \phi^\alpha \rangle \right] \tau^\beta \\ &\equiv e^\alpha ||\psi^{\alpha\beta}||^2 + (b_2^\alpha + Q^\alpha) \tau^\beta \end{aligned} \quad (4.6)$$

where Q^α is $2\pi/3$ times the sum of the electron densities at the nuclei in cluster α .

V. THE EXPECTATION VALUE OF V

For the evaluation of $\langle \psi | V | \psi \rangle$ we shall heavily use Newton's theorem according to which the potential of a spherically symmetric charge distribution with net charge zero vanishes outside the support of this charge distribution. This theorem makes the $\langle \Phi | V | \Phi \rangle$ term vanish upon rotation of the nuclei because it becomes the electrostatic interaction between two nonoverlapping spherical charge distributions of net charge zero. To demonstrate this formally we note that

$$\int d\mathcal{R} \langle \Phi(\underline{x}, \mathcal{R}\underline{X}) | V(\underline{x}, \mathcal{R}\underline{X}) | \Phi(\underline{x}, \mathcal{R}\underline{X}) \rangle$$

is a sum of terms

$$\int d\mathcal{R}^\alpha \int d\mathcal{R}^\beta \langle \phi^\alpha(\underline{x}^\alpha, \mathcal{R}^\alpha \underline{X}^\alpha) \phi^\beta(\underline{x}^\beta, \mathcal{R}^\beta \underline{X}^\beta) | V^{\alpha\beta}(\underline{x}^\alpha, \underline{x}^\beta, \mathcal{R}^\alpha \underline{X}^\alpha, \mathcal{R}^\beta \underline{X}^\beta, \mathbf{R}^{\alpha\beta}) | \phi^\alpha(\underline{x}^\alpha, \mathcal{R}^\alpha \underline{X}^\alpha) \phi^\beta(\underline{x}^\beta, \mathcal{R}^\beta \underline{X}^\beta) \rangle. \quad (5.1)$$

They involve only the one-electron density which upon integrating over $\mathcal{R}$ becomes spherically symmetric (around the center R^α) since $\phi^\alpha(\underline{x}^\alpha, \mathcal{R}^\alpha \underline{X}^\alpha) = \phi^\alpha((\mathcal{R}^\alpha)^{-1} \underline{x}^\alpha, \underline{X}^\alpha)$. The one-electron density is

$$\rho^\alpha(\mathbf{x}_1^\alpha) \equiv n^\alpha \int d\mathcal{R}^\alpha \int d\mathbf{x}_2^\alpha \cdots \int d\mathbf{x}_{n^\alpha}^\alpha | \phi^\alpha(\mathbf{x}_1^\alpha, \mathbf{x}_2^\alpha, \dots, \mathbf{x}_{n^\alpha}^\alpha, \mathcal{R}^\alpha \underline{X}^\alpha) |^2. \quad (5.2)$$

For $|\mathbf{x}| < R^\alpha - 1$, we have $\rho^\alpha(\mathbf{x}) = [1 + C \exp(-R^\alpha)] \rho_0^\alpha(\mathbf{x})$ for some $C > 0$, where $\rho_0^\alpha(\mathbf{x})$ is the "true" electron density calculated with ϕ_0^α .

With the averaged nuclear charge

$$n^\alpha(|\mathbf{x}|) = \int d\mathcal{R}^\alpha \sum_{j=1}^{M^\alpha} \delta(\mathbf{x} - \mathcal{R}^\alpha \mathbf{X}_j^\alpha), \quad (5.3)$$

and similar definitions for cluster β , (5.1) becomes

$$\int d\mathbf{x} \int d\mathbf{y} [\rho^\alpha(\mathbf{x}) - n^\alpha(\mathbf{x})] [\rho^\beta(\mathbf{y}) - n^\beta(\mathbf{y})] | \mathbf{x} - \mathbf{y} + \mathbf{R}^{\alpha\beta} |^{-1} = 0. \quad (5.4)$$

In the terms proportional to λ we have the factor $\phi^\alpha P^\alpha \phi^\alpha = \frac{1}{2} P^\alpha (\phi^\alpha)^2$. After partial integration, the gradient acts on the potential, and thus we obtain the dipole-dipole interaction directly:

$$\begin{aligned} &\int d\mathcal{R}^\alpha \int d\mathcal{R}^\beta \langle \phi^\alpha, \phi^\beta | V^{\alpha\beta} | \psi^\alpha, \psi^\beta \rangle \\ &= \int d\mathcal{R}^\alpha \int d\mathcal{R}^\beta \int d\underline{x}^\alpha \int d\underline{x}^\beta \frac{1}{4} \sum_i \mathbf{m}_{\alpha\beta} \cdot \nabla_i^\alpha | \phi^\alpha(\underline{x}^\alpha, \mathcal{R}^\alpha \underline{X}^\alpha) |^2 \sum_j \mathbf{n}_{\alpha\beta} \cdot \nabla_j^\beta | \phi^\beta(\underline{x}^\beta, \mathcal{R}^\beta \underline{X}^\beta) |^2 V(\underline{x}^\alpha, \underline{x}^\beta, \mathcal{R}^\alpha \underline{X}^\alpha, \mathcal{R}^\beta \underline{X}^\beta, \mathbf{R}^{\alpha\beta}) \\ &= \frac{1}{4} \int d\mathbf{x} \int d\mathbf{y} \rho^\alpha(\mathbf{x}) \rho^\beta(\mathbf{y}) (\mathbf{m}_{\alpha\beta} \cdot \nabla_{\mathbf{x}}) (\mathbf{n}_{\alpha\beta} \cdot \nabla_{\mathbf{y}}) | \mathbf{x} - \mathbf{y} + \mathbf{R}^{\alpha\beta} |^{-1} \\ &= \frac{1}{4} n^\alpha n^\beta [3(\mathbf{m}_{\alpha\beta} \cdot \mathbf{R}^{\alpha\beta})(\mathbf{n}_{\alpha\beta} \cdot \mathbf{R}^{\alpha\beta}) - (\mathbf{m}_{\alpha\beta} \cdot \mathbf{n}_{\alpha\beta}) | \mathbf{R}^{\alpha\beta} |^2] | \mathbf{R}^{\alpha\beta} |^{-5}. \end{aligned} \quad (5.5)$$

Newton's theorem has been used again in order to perform the dx and dy integrations. The contribution $\langle \psi^{\alpha\beta} | V^{\alpha\beta} | \psi^{\alpha\beta} \rangle$ is analogous to the term $\langle \phi | V^{\alpha\beta} | \phi \rangle$ except that the one-electron density is now calculated with $\psi^\alpha = \mathbf{m} \cdot \sum_{j=1}^n \nabla_j^2 \phi^\alpha$ instead of ϕ^α . If we call this density $\rho_m^\alpha(\mathbf{x})$ it will not be spherically symmetric, but $\rho_m^\alpha(\mathbf{x}) = \rho^\alpha(\mathbf{x})$. Since it is quadratic in $\mathbf{m}$ it must be of the form $|\mathbf{m}|^2 f_1(|\mathbf{x}|) + (\mathbf{m} \cdot \mathbf{x})^2 f_2(|\mathbf{x}|)$. Adding in the averaged charge of the nuclei will give a charge distribution $Q(x)$ with no net charge, and no dipole moment since $Q(x) = Q(-x)$. However, it may have a quadrupole moment and we may pick up a term

$$\langle \psi^{\alpha\beta} | V^{\alpha\beta} | \psi^{\alpha\beta} \rangle = c^{\alpha\beta} | \mathbf{R}^{\alpha\beta} |^{-5} \quad (5.6)$$

of unknown sign.

VI. CHOICE OF THE POLARIZATION DIRECTIONS

Our goal is to make the term proportional to λ , which will be negative for a suitable choice of λ , as big as possible while the term proportional to λ^2 should be kept small. Only (5.5) contributes linearly in λ and its maximal value $2n^{\alpha\beta} |\mathbf{R}^{\alpha\beta}|^{-3}$ is reached if we choose $\mathbf{m}_{\alpha\beta}$ and $\mathbf{n}_{\alpha\beta}$ in the direction of $\mathbf{R}^{\alpha\beta}$. However, this leaves us with the quadrupole term (5.6). We can get rid of the latter if we average over $\mathbf{m}_{\alpha\beta}$ and $\mathbf{n}_{\alpha\beta}$ in the following way. Let

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

be the unit vector in the direction of $\mathbf{R}^{\alpha\beta}$ and

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

two directions completing the orthogonal basis. Then average the polarizations over the following nine pairs of vectors:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix};$$

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix};$$

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

In this way the quadrupole term (5.6) vanishes since $Q(x)$ becomes spherically symmetric. On the other hand (5.5) averages to $\frac{1}{9} n^{\alpha\beta} |\mathbf{R}^{\alpha\beta}|^{-3}$. Since λ will be proportional to R^{-3} the quadrupole term (5.6) will contribute to the order R^{-11} and the first choice (i.e., no averaging) will give a better estimate for large R . On the other hand, for smaller distances averaging may be better, as (5.6) is not easy to estimate accurately.

VII. CHOICE OF THE $\lambda_{\alpha\beta}$

So far we have evaluated expressions such as $\langle \psi^{\alpha\beta} | V^{\alpha\beta} | \psi^{\alpha'\beta'} \rangle$ where all three index pairs $(\alpha\beta)$, $(\alpha'\beta')$, and $(\alpha''\beta'')$ are equal. The orthogonality (3.7) insures that terms in which they are not equal vanish. Similarly $\langle \psi^{\alpha\beta} | H^\gamma | \psi^{\alpha'\beta'} \rangle$ is not zero only if $(\alpha\beta) = (\alpha'\beta')$ and γ is either α or β . Thus collecting our results we arrive at

$$\begin{aligned} \sum_{\alpha} \int d\mathcal{R} \langle \psi | H | \psi \rangle &= \sum_{\alpha} \int d\mathcal{R} \left[\sum_{\alpha=1}^{\mathcal{C}} \left(e^{\alpha} |\psi|^2 + \sum b_1^{\alpha} \right) + \frac{1}{9} \sum_{\substack{\alpha, \beta \\ (1 \leq \alpha < \beta \leq \mathcal{C})}} \lambda_{\alpha\beta} n^{\alpha\beta} |\mathbf{R}^{\alpha\beta}|^{-3} \right. \\ &\quad \left. + \sum_{\substack{\alpha, \beta \\ (1 \leq \alpha < \beta \leq \mathcal{C})}} \lambda_{\alpha\beta}^2 [(b_2^{\alpha} + Q^{\alpha})\tau^{\beta} + (b_2^{\beta} + Q^{\beta})\tau^{\alpha}] \right]. \end{aligned} \quad (7.1)$$

To get an upper bound for the energy we have to extract a factor $|\psi|^2 = 1 + \sum_{\alpha < \beta} \lambda_{\alpha\beta}^2 \tau_{\alpha} \tau_{\beta}$ from the term in large square brackets. The minimization with respect to the $\lambda_{\alpha\beta}$'s then leads to a coupled system of cubic equations which cannot be solved analytically. For our purpose, however, it is sufficient to minimize the last two terms by putting

$$\lambda_{\alpha\beta} = -\frac{1}{18} n^{\alpha\beta} |\mathbf{R}^{\alpha\beta}|^{-3} [(Q^{\alpha} + b_2^{\alpha})\tau^{\beta} + (Q^{\beta} + b_2^{\beta})\tau^{\alpha}]^{-1}. \quad (7.2)$$

By extracting

$$|\psi|^2 = 1 + \frac{1}{(18)^2} \sum (n^{\alpha\beta})^2 \tau_{\alpha} \tau_{\beta} |\mathbf{R}_{\alpha\beta}|^{-6} [(Q^{\alpha} + b_2^{\alpha})\tau^{\beta} + (Q^{\beta} + b_2^{\beta})\tau^{\alpha}]^{-2},$$

we obtain

$$\sum_m \int d\mathcal{R} \langle \psi | H | \psi \rangle = \sum_m \int d\mathcal{R} ||\psi||^2 \left[\sum_{\alpha=1}^{\mathcal{C}} (e^\alpha + b_1^\alpha / ||\psi||^2) - \sum_{\substack{\alpha, \beta \\ (1 \leq \alpha < \beta \leq \mathcal{C})}} \frac{(n^\alpha n^\beta)^2}{(18)^2 [(Q^\alpha + b_1^\alpha) \tau^\beta + (Q^\beta + b_2^\beta) \tau^\alpha]} \right. \\ \left. \times \left[|\mathbf{R}^{\alpha\beta}|^6 + \sum_{\substack{\sigma, \rho \\ (1 \leq \sigma < \rho \leq \mathcal{C})}} \frac{(n^\sigma n^\rho)^2 \tau^\sigma \tau^\rho}{(18)^2 [(Q^\sigma + b_1^\sigma) \tau^\rho + (Q^\rho + b_2^\rho) \tau^\sigma]} \right]^{-1} \right]. \quad (7.3)$$

Taking $m_{\alpha\beta}$ and $n_{\alpha\beta}$ in the direction of $\mathbf{R}^{\alpha\beta}$ would change the factor $(18)^2$ into 4^2 and add $c^{\alpha\beta} |\mathbf{R}^{\alpha\beta}|^{-5}$ to $(Q^\alpha + b_1^\alpha) \tau^\beta + (Q^\beta + b_2^\beta) \tau^\alpha$.

VIII. SOME ESTIMATES

Our result (7.3) contains two parameters, the squared center-of-mass momentum τ , and Q , the density of the electrons at the nuclei. For the former we have the trivial estimate (where T^α is the kinetic energy in the ground state)

$$\tau^\alpha \leq n^\alpha \left\langle \phi^\alpha \left| -\sum_{k=1}^{n^\alpha} \Delta_k^\alpha \right| \phi^\alpha \right\rangle \equiv n^\alpha T^\alpha. \quad (8.1)$$

From the stability of matter one knows that for states of negative energy the kinetic energy is bounded by n^α ,¹⁴ so that $\tau^\alpha \leq c(n^\alpha)^2$. We conjecture that actually $\tau^\alpha \leq T^\alpha$. Similarly, for the Q^α the bound from Ref. 15 for the atomic case can be extended to the molecular case to give

$$Q^\alpha \leq \frac{1}{3} \sum_{k,j} \langle \phi^\alpha | Z_j^\alpha | x_k^\alpha - x_j^\alpha |^{-2} | \phi^\alpha \rangle. \quad (8.2)$$

Using the fact that $|x|^{-2} \leq -4\Delta$, the right side of (8.2) can be bounded by the kinetic energy. Again, keeping $\max_{\alpha,j} \{Z_j^\alpha\}$ fixed, we get a bound proportional to $(n^\alpha)^2$, whereas we conjecture it should be proportional to n^α . From the proven bounds the coefficient in front of the large parentheses in (7.3) becomes independent of the n^α . For the conjectured bounds it would be proportional to $n^\alpha n^\beta$, thereby indicating a linearity of the van der Waals forces with respect to the electron number in each cluster.

If one calculates our constants τ and Q for a hydrogen

atom one gets a constant in front of R^{-6} which is about an order of magnitude below the known constant for the van der Waals force between two hydrogen atoms. The reason is that our way of generating polarization by a rigid shift is a rather brutal act. In particular, shifting the electrons near the nucleus is energetically costly; it is much better to adjust only the outer parts of the electron cloud. The best way to do this will depend on the exact shape of the cluster, and a simple, general bound which is numerically good in all cases seems to be beyond reach. Our result shows that irrespective of these details one can always get some R^{-6} attraction by a rigid shift.

It will be noted that the factor in large parentheses in (7.3) contains a constant term, $\sum$, in addition to $|\mathbf{R}^{\alpha\beta}|^6$. While this term is asymptotically negligible for large $|\mathbf{R}^{\alpha\beta}|$, it unfortunately grows with $\mathcal{C}$. Our result is therefore useless if $\mathcal{C}$ is very large, e.g., in a crystal where $\mathcal{C} \sim 10^{23}$. To circumvent this difficulty a better trial function Ψ is needed. We believe that the following choice is adequate, but we have not actually pursued the matter: Define the "operator" $D^{\alpha\beta}$ by $D^{\alpha\beta} \phi^\alpha \phi^\beta = \psi^\alpha \psi^\beta$, $D^{\alpha\beta} \phi^\alpha \psi^\beta = D^{\alpha\beta} \psi^\alpha \phi^\beta = D^{\alpha\beta} \psi^\alpha \psi^\beta = 0$. Then a natural generalization of (3.9) is

$$\Psi = \prod_{\substack{\alpha, \beta \\ (\alpha < \beta)}} (1 + \lambda_{\alpha\beta} D^{\alpha\beta}) \prod_{\alpha=1}^{\mathcal{C}} \phi^\alpha(\underline{x}^\alpha, \underline{X}^\alpha). \quad (8.3)$$

ACKNOWLEDGMENTS

This work was partially supported by U.S. National Science Foundation Grant No. PHY-81-16101-A03.

¹P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

²E. H. Lieb, Int. J. Quantum Chem. **24**, 243 (1983).

³E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981); **54**, 311(E) (1982).

⁴B. R. Junker and J. N. Bardsley, Phys. Rev. Lett. **28**, 1227 (1972).

⁵D. L. Morgan, Jr. and V. W. Hughes, Phys. Rev. A **7**, 1811 (1973); Phys. Rev. D **2**, 1389 (1970).

⁶W. Kolos, D. L. Morgan, D. M. Schrader, and L. Wolniewicz, Phys. Rev. A **6**, 1792 (1975).

⁷G. Feinberg and J. Sucher, Phys. Rev. A **2**, 2395 (1970).

⁸H. Margenau and N. R. Kestner, *The Theory of Intermolecular Forces* (Pergamon, New York, 1971).

⁹Yu. S. Barash and V. L. Ginzburg, Usp. Fiz. Nauk **143**, 345 (1984) [Sov. Phys.—Usp. **27**, 7 (1984)].

¹⁰B. Simon and J. Morgan, Int. J. Quantum Chem. **17**, 1143 (1980).

¹¹H. G. Casimir and D. Polder, Phys. Rev. **73**, 360 (1948).

¹²E. M. Lifshitz, Zh. Eksp. Teor. Fiz. **29**, 94 (1955).

¹³J. R. Manson and R. H. Ritchie, Phys. Rev. Lett. **54**, 785 (1985).

¹⁴E. H. Lieb, Rev. Mod. Phys. **48**, 553 (1976).

¹⁵M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and W. Thirring, J. Phys. B **11**, L571 (1978).

Electron density near the nucleus of a large atom

Ole J. Heilmann

Chemistry Laboratory 3, H.C. Orsted Institute, University of Copenhagen, Universitetsparken 5, DK2100 Copenhagen, Denmark

Elliott H. Lieb

Departments of Physics and Mathematics, Princeton University, P.O. Box 708, Princeton, New Jersey 08544-0708

(Received 1 May 1995)

The density of electrons on a distance scale $1/Z$ near the nucleus of a large atom with nuclear charge Ze is given (asymptotically as $Z \rightarrow \infty$) by the sum of the squares of all the hydrogenic bound-state functions (with nuclear charge Ze). This density function, which is an important limiting function in quantum chemistry, is investigated here in detail. Several analytic results are found: In particular, the asymptotic expansion for large r is derived and it is shown that the function falls off as $r^{-3/2}$ for large r ; this behavior coincides with the Thomas-Fermi density for small r . “Shell structure” is visible, but barely so.

PACS number(s): 31.10.+z, 31.15.An, 31.15.Ew

I. INTRODUCTION AND BASIC DEFINITIONS

In attempting to understand the ground state of an atom (for the nonrelativistic Schrödinger equation with an infinitely massive fixed nucleus) it is useful to explore the properties of this ground state as the nuclear charge Ze tends to infinity. In this study a function of fundamental importance arises, and it is our purpose here to explore its properties. That function is the sum of the squares of all the bound-state eigenfunctions of the hydrogen atom.

The *Schrödinger Hamiltonian* for an atom with N electrons is

$$H_{N,Z} = -\frac{\hbar^2}{2m} \sum_{i=1}^N \Delta_i - Ze^2 \sum_{i=1}^N |x_i|^{-1} + e^2 \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}. \quad (1.1)$$

Here $-e$ and m are, respectively, the charge and mass of an electron; the x_i 's are the electron coordinates and Δ is the three-dimensional Laplacian. The first operator in (1.1) is the kinetic energy, the second is the attraction of the electrons to the nucleus (located at the origin in $\mathbb{R}^3$), and the third term is the electron-electron repulsion.

The ground-state energy $E_{N,Z}$ is defined to be

$$E_{N,Z} = \inf_{\psi} \langle \psi | H_{N,Z} | \psi \rangle, \quad (1.2)$$

where the infimum is taken over all functions $\psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)$ of N space and spin variables (x_i, σ_i) that are normalized, i.e.,

$$\sum_{\sigma_1=1}^2 \cdots \sum_{\sigma_N=1}^2 \int |\psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 \times d^3x_1 \cdots d^3x_N = 1, \quad (1.3)$$

and that are antisymmetric [i.e., ψ changes sign if (x_i, σ_i) is interchanged with (x_j, σ_j) for any $i \neq j$]. For any normalized ψ (not necessarily a ground state ψ), there is the *single particle density* ρ_{ψ} defined by

$$\rho_{\psi}(x) = N \sum_{\sigma_1=1}^2 \cdots \sum_{\sigma_N=1}^2 \int |\psi(x, x_2, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 \times d^3x_2 \cdots d^3x_N, \quad (1.4)$$

which satisfies

$$\int \rho_{\psi}(x) d^3x = N. \quad (1.5)$$

Note that because of the antisymmetry of ψ it does not matter which x_i is set equal to x in (1.4).

If $N \leq Z$, then [1] there is at least one square integrable ψ that actually minimizes the right-hand side of (1.2). Such a ψ is a *ground state* and satisfies Schrödinger's equation

$$H_{N,Z}\psi = E_{N,Z}\psi. \quad (1.6)$$

Of course a ground state might exist when $N > Z$ (i.e., for a negative ion), but it is not known how large N can be (see [2] for a review). In particular it is known [2] that N must satisfy $N < 2Z + 1$ and [3] that as $Z \rightarrow \infty$ the maximum N satisfies $N/Z \rightarrow 1$. In fact [4], $|N/Z - 1| < Z^{-1/7}$.

Our goal here is to study $\rho_{\psi}(x)$ for a ground state ψ when Z is large. To be precise, fix some number λ with $0 < \lambda \leq 1$, let N take the values $N = 1, 2, 3, \dots$, and for each N let Z_N be defined by $\lambda Z_N = N$. For each integer N let $\psi_{N,Z}$ be a ground state and let $\rho_{N,Z}$ be given by (1.4) with ψ replaced by $\psi_{N,Z}$. Of course $\psi_{N,Z}$ (and hence $\rho_{N,Z}$) may not be unique, but nevertheless the following is true [5,6], for any choice of the $\psi_{N,Z}$'s: As N and Z tend to infinity, with $N = \lambda Z_N$ and with λ fixed

$$Z_N^{-2} \rho_{N,Z}(Z_N^{-1/3}x) \rightarrow \rho_{\lambda}^{\text{TF}}(x). \quad (1.7)$$

Here ρ_λ^{TF} is the Thomas-Fermi (TF) density. It is the unique non-negative function that minimizes the TF energy functional for a unit nuclear charge,

$$\mathcal{E}^{\text{TF}}(\rho) = \frac{3}{5} \frac{\hbar^2}{2m} (3\pi^2)^{2/3} \int \rho(x)^{5/3} d^3x - e^2 \int \rho(x) \frac{1}{|x|} d^3x + \frac{1}{2} e^2 \iint \rho(x)\rho(y) \frac{1}{|x-y|} d^3x d^3y, \quad (1.8)$$

under the condition that

$$\int \rho(x) d^3x = \lambda. \quad (1.9)$$

The unique minimizing $\rho_\lambda^{\text{TF}}(x)$ is the unique solution to the TF equation

$$\frac{\hbar^2}{2m} (3\pi^2)^{2/3} [\rho_\lambda^{\text{TF}}(x)]^{2/3} = e^2 \left[\frac{1}{|x|} - \int \frac{1}{|x-y|} \rho_\lambda^{\text{TF}}(y) d^3y - \mu \right]_+, \quad (1.10)$$

where $[t]_+$ equals t if $t > 0$ and 0 if $t \leq 0$. In (1.10), μ is the chemical potential [given by $-dE^{\text{TF}}(\lambda)/d\lambda$] and $E^{\text{TF}}(\lambda)$ is the TF energy given by

$$E^{\text{TF}}(\lambda) = \mathcal{E}^{\text{TF}}(\rho_\lambda). \quad (1.11)$$

The convergence as $N \rightarrow \infty$ in (1.7) occurs in several senses. First, there is the energy convergence

$$Z^{-7/3} E_{N,Z} \rightarrow E^{\text{TF}}(\lambda) \quad \text{as } N \rightarrow \infty, \quad (1.12)$$

where $N/\lambda = Z$ as before. This means that TF theory exactly gives the leading term in the ground-state energy. Second, $Z_N^{-2} \rho_{N,Z}(Z^{-1/3}x)$ converges to $\rho_\lambda^{\text{TF}}(x)$ in the following sense as $N \rightarrow \infty$:

$$\int_A [Z_N^{-2} \rho_{N,Z}(Z^{-1/3}x) - \rho_\lambda^{\text{TF}}(x)] d^3x \rightarrow 0 \quad (1.13)$$

for any measurable set A in $\mathbb{R}^3$. Note that for each N

$$\int Z_N^{-2} \rho_{N,Z}(Z_N^{-1/3}x) d^3x = \frac{N}{Z_N} = \lambda = \int \rho_\lambda^{\text{TF}}(x) d^3x, \quad (1.14)$$

which means that the TF density correctly accounts for *all* of the electrons. Equation (1.13) implies that “most of the electrons” in a large atom are at a distance of order $Z^{-1/3}$ (or $N^{-1/3}$) from the nucleus. Thus, in one sense at least, a large atom has a smaller radius than a small atom. This radius $Z^{-1/3}$ answers the question, where are most of the electrons, but it is not the *chemical radius*, which can be defined as the radius beyond which there is one electron. This chemical radius is presumably of the order unity (i.e., independent of N), but no one has succeeded in proving this so far.

There is, however, another important radius, namely, Z^{-1} . This is the radius of the innermost, or K -shell, electrons. Quantum effects at this inner radius give rise to the next correction to the ground-state energy, i.e.,

$$E_{N,Z} = Z^{7/3} E^{\text{TF}}(\lambda) + Z^2/4 + (\text{lower order terms}). \quad (1.15)$$

The second term $Z^2/4$ is the correction predicted by Scott [7], which has been proved to be correct by Siedentop and Weikard [8,9] and Hughes [10]. It is important to note that this second term is *independent* of λ and is therefore consistent with the view that as long as N is of the order Z (i.e., $\lambda > 0$) the innermost shells are always totally filled and the contributions of these shells (of radius of order Z^{-1}) to the energy and density are not affected by the bulk of the electrons, which are much farther out at a radius of order $Z^{-1/3}$.

How many electrons are in the inner shells? Thomas-Fermi theory would predict [from (1.7) and (1.10)] that when $|x| \ll Z^{-1/3}$ and $N = \lambda Z$

$$\rho_{N,Z}(x) \approx (2m/\hbar^2)^{3/2} (3\pi^2)^{-1} (Z/|x|)^{3/2}. \quad (1.16)$$

Note that this form, (1.16), depends only on Z and not on N and therefore is consistent with the remark in the preceding paragraph. When $|x| \approx Z^{-1}$, then (1.16) predicts that $\rho_{N,Z}(x) \approx Z^3$, which is reasonable, but as $|x| \rightarrow 0$ the right-hand side of (1.16) tends to infinity and this cannot be a correct conclusion. In other words, the TF prediction (1.16) has to be modified for $|x|$ of the order Z^{-1} or smaller and it is this modification that ultimately leads to the Scott correction to the energy $Z^2/4$.

There was a conjecture [6] about the appropriate corrections to (1.16) when $|x|$ is small. This is the *strengthened form of the Scott correction* and recently it has been proved [11]. The purpose of this paper is to try to elucidate the consequences of the strengthened form of the Scott correction, which is the following. For distances that are small compared to the Thomas-Fermi scale $Z^{-1/3}$, the density $\rho_N(x)$ is given for large N and Z by

$$\rho_{N,Z}(x) \approx \rho_Z^B(x), \quad (1.17)$$

where ρ_Z^B is the density in an infinite “Bohr atom” of nuclear charge Z , which depends on Z , but not on N . By this we mean the density in an atom with infinitely many electrons in which the electron-electron repulsion [i.e., the third term in (1.1) is omitted]. The precise statement of (1.17) is given in (1.21) below. This density ρ_Z^B can be calculated explicitly and we do so in the following. While the formal definition of ρ_Z^B is easy to obtain, the actual elucidation of the properties of this function, which is the purpose of this paper, turns out to be surprisingly difficult.

A. Definition of the hydrogenic density

Let $\psi_{n/lm}(x)$, with x in $\mathbb{R}^3$, denote the normalized bound-state Bohr orbitals of a hydrogen atom (with unit nuclear charge) and in atomic units in which $\hbar^2/m = 1$ and $e = 1$, i.e.,

$$\left[-\frac{1}{2} \Delta - \frac{1}{|x|} \right] \psi_{nlm} = e_n \psi_{nlm}. \quad (1.18)$$

Here n is the principal quantum number, l is the angular momentum, and m is the azimuthal quantum number. The energy is $e_n = -1/(2n^2)$. Then

$$\rho^H(x) = \sum_{n=1}^{\infty} \sum_{l=0}^{n-1} \sum_{m=-l}^l |\psi_{nlm}(x)|^2 \quad (1.19)$$

is the sum of the square of all these bound-state eigenfunctions. This sum, as will be seen later, is finite for all x . Note also that, since we sum on m in (1.19), $\rho^H(x)$ is a function only of the radius $|x| = r$. We shall abuse notation by writing $\rho^H(r)$ instead of $\rho^H(x)$.

The relation between ρ_Z^B and ρ^H is

$$\rho_Z^B(x) = 2 \left[\frac{Ze^2m}{\hbar^2} \right]^3 \rho^H \left(\frac{Ze^2m}{\hbar^2} x \right), \quad (1.20)$$

in which the factor 2 arises from the two spin states of the electron. The precise statement of (1.17) is that as N and Z tend to infinity with $N/Z = \lambda$,

$$\left[\frac{\hbar^2}{Z_N e^2 m} \right]^3 \rho_{N,Z} \left(\frac{\hbar^2}{Z_N e^2 m} x \right) - 2 \rho^H(x) \rightarrow 0 \quad (1.21)$$

in two senses. One is the three-dimensional integral sense as in (1.13). The second and stronger sense is pointwise (with respect to radius) on spheres: If S_R is a sphere of radius $R > 0$ centered at the origin and if A is any (two-dimensional) measurable subset of S_R , then the (two-dimensional) integral over A of the left-hand side of (1.21) converges to zero. This is proved in [11].

B. Definition of ψ_{nlm}

In polar coordinates $x = r (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$,

$$\psi_{nlm}(x) = R_{nl}(r) Y_{lm}(\theta, \phi). \quad (1.22)$$

Here Y_{lm} is a normalized spherical harmonic

$$\int_0^{2\pi} \int_0^\pi |Y_{lm}(\theta, \phi)|^2 \sin \theta \, d\theta \, d\phi = 1. \quad (1.23)$$

The radial function is

$$R_{nl}(r) = N_{nl} (2r/n)^l e^{-r/n} L_{n-l-1}^{2l+1}(2r/n). \quad (1.24)$$

The normalization constant N_{nl} is given by

$$(N_{nl})^2 = 4n^{-4} \frac{(n-l-1)!}{(n+l)!} \quad (1.25)$$

and is chosen so that

$$\int_0^\infty R_{nl}(r)^2 r^2 dr = 1. \quad (1.26)$$

The function $\dot{L}_n^m$ is the Laguerre polynomial (see [12], Sec. 10.12)

$$L_n^m(z) = \sum_{j=0}^n \binom{n+m}{n-j} \frac{1}{j!} (-z)^j. \quad (1.27)$$

The summation on m in (1.19) can be done easily:

$$\sum_{m=-l}^l |Y_{lm}(\theta, \phi)|^2 = \frac{1}{4\pi} (2l+1); \quad (1.28)$$

therefore

$$\rho^H(r) = \frac{1}{4\pi} \sum_{n=1}^{\infty} \sum_{l=0}^{n-1} (2l+1) [R_{nl}(r)^2]. \quad (1.29)$$

The simplest case of R_{nl} is $l = n-1$, where

$$R_{n,n-1}(r) = 2n^{-2} [(2n-1)!]^{-1/2} (2r/n)^{n-1} e^{-r/n}, \quad (1.30)$$

which is easily seen to satisfy the normalization condition (1.26) and the Schrödinger equation (1.18) with $e_n = 1/(2n^2)$.

II. ANALYTIC PROPERTIES OF THE HYDROGENIC DENSITY ρ^H

A. Large distance asymptotics

The function $\rho^H(x)$ is a fixed function, but the right-hand side of (1.20) defines a function that has a scale length Z^{-1} and a typical amplitude Z^3 . The Thomas-Fermi prediction (1.7), namely, $Z^2 \rho_\lambda^{\text{TF}}(Z^{1/2}x)$, is a function with scale length $Z^{-1/3}$ and a typical amplitude Z^2 . Since the strong form of the Scott correction is correct these two functions must come together smoothly. This means that the *large distance* behavior of $\rho^H(x)$ must coincide with the *short distance* behavior of $\rho^{\text{TF}}(x)$, which is given by (1.16) with $\hbar/m = Z = 1$. This is indeed true.

Theorem 1 (large distance behavior of ρ^H). As $r \rightarrow \infty$, $\rho^H(r)$ has the asymptotic expansion

$$\rho^H(r) = \frac{1}{\pi^2 \sqrt{2}} r^{-3/2} \left[\sum_{j=0}^{\infty} a_j (8r)^{-j} - \sin(\sqrt{32}r) \sum_{j=1}^{\infty} b_j (8r)^{-j} + \cos(\sqrt{32}r) \sum_{j=1}^{\infty} c_j (8r)^{-j-1/2} \right]. \quad (2.1)$$

The first few coefficients are

$$a_0 = 2/3, \quad a_1 = -1/12, \quad a_2 = 79/960,$$

$$b_1 = 3/2, \quad b_2 = -140 \, 589/11 \, 200,$$

$$c_1 = 141/40, \quad c_2 = -2 \, 028 \, 627/44 \, 800.$$

In particular,

$$\rho^H(r) = 2^{1/2} (3\pi^2)^{-1} r^{-3/2} + O(r^{-5/2}), \quad (2.2)$$

which is equivalent to

$$\lim_{r \rightarrow \infty} r^{3/2} \rho^B(r) = \lim_{r \rightarrow 0} r^{3/2} \rho_\lambda^{\text{TF}}(r) \quad (2.3)$$

for each $0 < \lambda \leq 1$.

Equation (2.3) is thus consistent with (1.16), (1.17), (1.20), and (1.21). The truth of Eq. (2.3) is motivated by the physical considerations at the beginning of this section. In Appendix F we derive the asymptotic expansion, the coefficients of which can, in principle, be calculated to arbitrarily high order.

Two important facts should be noted about this theorem. The first is that the oscillations of $\rho^H(r)$ (which is often called the “shell structure”) are quite small. They hardly leap to the eye in the graph of $\rho^H(r)$ (Fig. 1). The second fact is that the period of the oscillations (as a function of $\sqrt{r}$, not r) is $2\pi/\sqrt{32}=1.11$. This is what one would expect from the fact that the average potential energy $-1/r$ for principle quantum number n is $2e_n=n^{-2}$. The fact that $\sqrt{r}$ does not peak at n but rather at slightly larger value $1.11n$ is consistent with Hölder’s inequality applied to expectation values $\langle \cdot \rangle_n$ at the n th level; this inequality is, in general, $\langle r^{1/2} \rangle \langle r^{-1} \rangle^{1/2} \geq 1$. What is a bit unexpected is that the increase (from 1 to 1.11) is independent of n (see Fig. 3). The following is important for the proof of Theorem 1.

Lemma 1 (first integral representation).

$$\rho^H(r) = \pi^{-2}(2r)^{-3/2} \int_0^\infty x e^{-x} \phi(x) \int_0^\pi w(x, \theta) \times J_3(2\phi(x)[2r w(x, \theta)]^{1/2}) d\theta dx, \quad (2.4)$$

where J_3 is a Bessel function and

$$\phi(x) = [x/(1 - e^{-x})]^{1/2} \quad (2.5)$$

and

$$w(x, \theta) = 1 + e^{-x} - 2e^{-x/2} \cos \theta. \quad (2.6)$$

Part of the proof of Lemma 1, which is given in Appendix B, rests on the fact that we are able to do the sum on l in (1.29) in closed form as follows.

Lemma 2 (the sum on a row). For all complex numbers z

$$S_n(z) := \sum_{l=0}^{n-1} (2l+1) \frac{(n-l-1)!}{(n+l)!} z^{2l} [L_{n-l-1}^{2l+1}(z)]^2 \quad (2.7)$$

satisfies

$$\begin{aligned} S_n(z) &= \sum_{m=0}^{n-1} (m!)^{-2} z^{2m} L_{n-m-1}^{2m+1}(2z) \\ &= n[L_{n-1}^1(z)]^2 + n[L_{n-2}^1(z)]^2 \\ &\quad + (z-2n)L_{n-1}^1(z)L_{n-2}^1(z). \end{aligned} \quad (2.8)$$

Moreover,

$$\frac{d}{dz} e^{-z} S_n(z) = -e^{-z} [L_{n-1}^1(z)]^2, \quad (2.9)$$

which implies that $e^{-z} S_n(z)$ is monotonically decreasing in z for real z .

Using the definitions (1.24) and (1.29) we have

$$\rho^H(r) = \frac{1}{\pi} \sum_{n=1}^{\infty} n^{-4} e^{-2r/n} S_n(2r/n), \quad (2.10)$$

$$\begin{aligned} \rho^H(r) &= \frac{1}{\pi} \sum_{n=1}^{\infty} n^{-4} e^{-2r/n} [n L_{n-1}^1(2r/n)^2 + n L_{n-2}^1(2r/n)^2 \\ &\quad + 2(r/n - n) L_{n-1}^1(2r/n) L_{n-2}^1(2r/n)] \end{aligned} \quad (2.11)$$

and

$$\frac{d}{dr} \rho^H(r) = -\frac{2}{\pi} \sum_{n=1}^{\infty} n^{-5} e^{-2r/n} [L_{n-1}^1(2r/n)]^2, \quad (2.12)$$

which shows that $\rho^H(r)$ is monotonically decreasing in r and hence that $\rho^H(r)$ is finite for all r and achieves its maximum at $r=0$:

$$\rho^H(r) \leq \rho^H(0) = \frac{1}{\pi} \sum_{n=1}^{\infty} n^{-3} \approx 0.383. \quad (2.13)$$

Lemma 2 will be proved in Appendix A. It will be useful for us to recast Lemma 1 in the following form, which will be proved in Appendix C.

Lemma 3 (second integral representation). The density ρ^H can be written as

$$\begin{aligned} \rho^H(r) &= \frac{2}{\pi^2 (2r)^{3/2}} \int_0^\infty J_3(y \sqrt{8r}) y^{-2} \int_{\phi_3(y)}^{\psi_1(y)} [\phi_3(x)]^2 \\ &\quad \times [y^2 - \phi_1(x)^2]^{-1/2} [\phi_2(x) - y^2]^{-1/2} x dx dy, \end{aligned} \quad (2.14)$$

where

$$\phi_1(x) = [x \tanh(x/4)]^{1/2}, \quad (2.15)$$

$$\phi_2(x) = [x \coth(x/4)]^{1/2}, \quad (2.16)$$

$$\phi_3(x) = \frac{x}{2 \sinh(x/2)}, \quad (2.17)$$

$$\psi_3(y) = \begin{cases} 0 & \text{if } 0 \leq y \leq 2 \\ \psi_2(y) & \text{if } 2 \leq y, \end{cases}$$

and where ψ_1 and ψ_2 are the inverse functions of ϕ_1 and ϕ_2 , e.g., $\psi_i(y) = x$ when $y = \phi_i(x)$, $i=1,2$.

The second integral representation brings us to the following problem, which can be stated generally as follows. Find an asymptotic expansion, for large x , of integrals of Bessel functions of the form

$$\int_a^b J_\nu(xt) f(t) dt. \quad (2.18)$$

We formulate our results as two lemmas, which are proved in Appendix D and E, respectively, and we emphasize that our proofs are an adaptation of Olver’s general method [13].

Lemma 4 [asymptotics of (2.18) with $a=0$]. Let $f(t)$ have the expansion

$$f(t) \approx t^{\mu-\nu-1} \sum_{j=0}^{\infty} a_j t^j, \quad \mu > 0 \quad (2.19)$$

in the limit $t \rightarrow 0+$, valid in the sense that (2.19) represents both $f(t)$ and all its derivatives $f^{(n)}(t)$ asymptotically correctly in the limit $t \rightarrow 0+$. Assume further that $f(t)$ is continuously differentiable infinitely many times in the open interval $0 < t < b$ and that all the limits

$$\lim_{t \uparrow b} f^{(n)}(t) =: f^{(n)}(b), \quad n = 0, 1, 2, \dots \quad (2.20)$$

exist and are finite. Then we have the asymptotic expansion for $x \rightarrow +\infty$

$$\begin{aligned} \int_0^b J_\nu(xt) f(t) dt &= \sum_{j=0}^{\infty} a_j x^{\nu-\mu-j} 2^{\mu+j-\nu-1} \\ &\quad \times \Gamma((\mu+j)/2) / \Gamma(\nu+1-(\mu+j)/2) \\ &\quad + G_+(1, b^-, x), \end{aligned} \quad (2.21)$$

where

$$\begin{aligned} G_\pm(\lambda, b, x) &= \left(\frac{2}{\pi} \right)^{1/2} \sum_{s=0}^{\infty} x^{-s-\lambda-1/2} b^{-s-1/2} \\ &\quad \times \cos \left[x b - \left(\nu + s \pm \frac{1}{2} \right) \frac{\pi}{2} \right] \\ &\quad \times \sum_{j=0}^s \frac{f^{(j)}(b)}{j!} b^j \sum_{l=0}^{s-j} (-1)^{j+l} \frac{\Gamma(s-l+\lambda)}{l!(s-j-l)!} \\ &\quad \times \frac{\Gamma(s-j+\frac{1}{2}) \Gamma(\nu+l+\frac{1}{2})}{\Gamma(l+\frac{1}{2})! \Gamma(\nu-l+\frac{1}{2})} 2^{-l}. \end{aligned} \quad (2.22)$$

Lemma 5 [asymptotics of (2.18) with $a > 0$]. Let $f(t)$ have the expansion

$$f(t) \approx \sum_{j=0}^{\infty} a_j (t-a)^{j+\lambda-1}, \quad \lambda > 0 \quad (2.23)$$

in the limit $t \rightarrow a+$, valid in the sense that (2.23) represents both $f(t)$ and all its derivatives asymptotically correctly in the limit $t \rightarrow a+$. Assume further that $f(t)$ is continuously differentiable infinitely many times in the open interval $a < t < b$ and that all the limits

$$\lim_{t \uparrow b} f^{(n)}(t) =: f^{(n)}(b), \quad n = 0, 1, 2, \dots \quad (2.24)$$

exist and are finite. Then we have the asymptotic expansion for $x \rightarrow +\infty$

$$\int_a^b J_\nu(xt) f(t) dt = G_-(\lambda, a+, x) + G_+(1, b-, x), \quad (2.25)$$

where the functions $G_\pm$ are given by Eq. (2.22).

If a and b are interchanged (i.e., $b < a$), then the condition in the limit $t \rightarrow a-$ becomes

$$f(t) \approx \sum_{j=0}^{\infty} (-1)^j a_j (a-t)^{j+\lambda-1}, \quad \lambda > 0 \quad (2.26)$$

and the result becomes

$$\int_b^a J_\nu(xt) f(t) dt = G_-(1, b+, x) + G_+(\lambda, a-, x) \quad (2.27)$$

If we take the limit $b \rightarrow \infty$ in (2.25), we get the asymptotic expansion

$$\int_a^\infty J_\nu(xt) f(t) dt = G_-(\lambda, a+, x) \quad (2.28)$$

for $x \rightarrow +\infty$, provided $f(t)$ is continuously differentiable infinitely many times in the open interval $a < t < \infty$ and Eq. (2.23) still holds in the limit $t \rightarrow a+$, while the requirements in the limit $t \rightarrow +\infty$ are absolute integrability of $f(t)$ and

$$\lim_{t \rightarrow \infty} t^{-1/2} f^{(n)}(t) = 0, \quad n = 1, 2, \dots \quad (2.29)$$

We conclude with a heuristic derivation of the origin of the oscillatory terms in Theorem 2.1.

B. Large n asymptotics of $S_n(r)$

According to (2.9) the function of interest is

$$W_n(z) = e^{-z/2} L_{n-1}^1(z) \quad (2.30)$$

(but recall that we are interested in this for $z = 2r/n$). According to Erdelyi *et al.* [12], there is an asymptotic formula due to Tricomi that is useful for numerical computations:

$$W_n(z) = n^{1/2} z^{-1/2} \sum_{m=0}^{\infty} A_m (z/4n)^{m/2} J_{m+1}(\sqrt{4nz}), \quad (2.31)$$

in which J_{m+1} is a Bessel function and

$$A_0 = 1, \quad A_1 = 0, \quad A_2 = 1,$$

$$A_m = A_{m-2} - (2n/m) A_{m-3}, \quad m > 2. \quad (2.32)$$

This series is valid for $z < \text{const} \times n^{1/3}$.

The Bessel functions $J_{m+1}(x)$ are oscillatory for large x with period 2π and it is this oscillation that gives rise to the oscillation shown in Theorem 1. For large n and fixed $z > 0$, we have from (2.31) that [Fejér's formula, see [12], Eq. (10.15.1)]

$$W_n(z) = \pi^{-1/2} z^{-3/4} n^{1/4} \cos[2(nz)^{1/2} - 3\pi/4] + O(n^{-1/4}). \quad (2.33)$$

By substituting $z = 2r/n$ in (2.31) and then squaring W_n (which halves the period) we would conclude that the correction term in $\rho^H(r)$ has period $\pi/\sqrt{8}$ (in $\sqrt{r}$), as given in Theorem 1. This argument requires additional justification, however, for the following reason. We require $W_n(2r/n)$ for n of the order $\sqrt{r}$, which is much smaller than r when r is large. This means that $z = 2r/n$ is not always small and therefore the asymptotic formula (2.31) is not necessarily valid in the region of interest. Nevertheless, the conclusion just stated is correct. The oscillation period in $\sqrt{r}$ is indeed $\pi/\sqrt{8}$ as

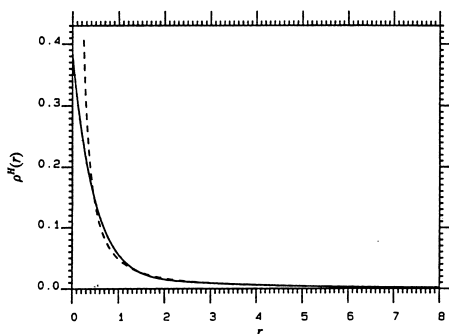


FIG. 1. $\rho^H(r)$ as a function of r . Also shown (dashed curve) is the Thomas-Fermi asymptotic expression $2^{1/2}(3\pi^2)^{-1}r^{-3/2}$.

(2.31) would seem to imply, but we do *not* assert that the amplitudes A_m are the correct ones. The correct formula is in Theorem 1.

III. SUMMARY OF RESULTS

We have analyzed the function $\rho^H(r)$, which is the sum of the squares of the hydrogenic bound-state functions. This function is the density that would be seen in a very large Z atom on a distance scale Z^{-1} from the nucleus. It is very nicely behaved, being finite everywhere and monotonically decreasing. We have derived integral representations for this function that are useful for analytical and numerical analysis. We have analyzed the large r behavior of this function in detail and discovered that, contrary to our expectation at least, there is very little oscillatory structure. In fact, it is necessary to take two derivatives with respect to r in order to make the oscillatory terms as large as the nonoscillatory ones. In short, shell structure is not a prominent property of this universal atomic function.

In order to clarify the nature of $\rho^H(r)$ we give a graph of this function together with its asymptotic Thomas-Fermi

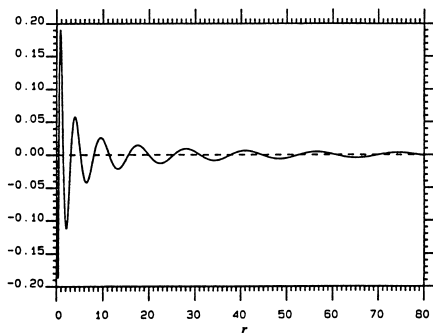


FIG. 2. Relative difference $\rho^H(r)/[2^{1/2}(3\pi^2)^{-1}r^{-3/2}] - 1$ as a function of r .

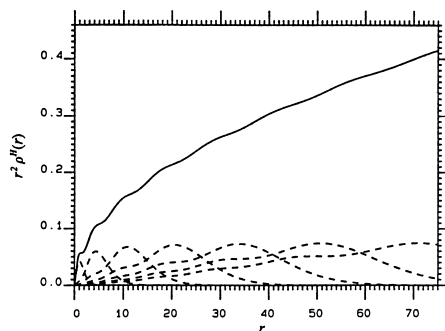


FIG. 3. $r^2 \rho^H(r)$ as a function of r . Also shown (dashed curves) are $r^2 S_n(2r/n)n^{-4}e^{-2r/n}/\pi$ for $n = 1 - 7$.

form $2^{1/2}(3\pi^2)^{-1}r^{-3/2}$ in Fig. 1. The difference between the two functions, which is the part of the density that is not semiclassical, is also plotted in Fig. 2. The conventional function $r^2 \rho^H(r)$ is plotted in Fig. 3.

Note: Appendices A-F, which contain the proofs of Lemmas 1-5 and Theorem 1, have been omitted in this reproduction. They can be found in the original publication.

- [1] G. Zhislin, Tr. Mosk. Mat. Obsch. **9**, 81 (1960).
- [2] E. H. Lieb, Phys. Rev. A **29**, 3018 (1984).
- [3] E. H. Lieb, I. M. Sigal, B. Simon, and W. Thirring, Phys. Rev. Lett. **52**, 994 (1980); Commun. Math. Phys. **116**, 635 (1988).
- [4] C. L. Fefferman and L. A. Seco, Commun. Math. Phys. **128**, 109 (1990).
- [5] E. H. Lieb and B. Simon, Adv. Math. **23**, 22 (1977).
- [6] E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981); **54**, 311(E) (1982).
- [7] J. M. C. Scott, Philos. Mag. **43**, 859 (1952).
- [8] H. Siedentop and R. Weikard, Invent. Math. **97**, 213 (1989).
- [9] H. Siedentop and R. Weikard, Commun. Math. Phys. **112**, 471 (1987).
- [10] W. Hughes, Adv. Math. **79**, 213 (1990).
- [11] A. Iantchenko, E. H. Lieb, and H. Siedentop, J. Reine Angew. Math. (to be published).
- [12] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
- [13] F. W. J. Olver, SIAM J. Math. **5**, 19 (1974).
- [14] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1962).

Proof of a conjecture about atomic and molecular cores related to Scott's correction

By *Alexei Iantchenko* at Rennes, *Elliott H. Lieb*¹⁾ at Princeton and *Heinz Siedentop*²⁾ at Oslo

1. Introduction

A great deal is known about the ground states of large atoms in the framework of the non-relativistic Schrödinger equation, with fixed (i.e., infinitely massive) nuclei. The leading term, in powers of the nuclear charge Z , is given exactly by Thomas-Fermi theory, as was proved by Lieb and Simon [12]; see [11] for a review. This leading term in the energy is proportional to $Z^{7/3}$, with the proportionality constant depending on the ratio of N/Z , which is assumed to be held fixed as $Z \rightarrow \infty$. Here, N is the electron number. Neutrality, i.e., $N = Z$, is not required, even though it is the case of primary physical interest. The characteristic length scale for the electron density (in the sense that all the electrons can be found on this scale in the limit $Z \rightarrow \infty$) is $Z^{-1/3}$. The fact that the true quantum-mechanical electron density, ρ_d , converges (after suitable scaling) to the Thomas-Fermi density, ρ^{TF} , as $Z \rightarrow \infty$ with N/Z fixed was proved in [12]. The chemical radius, which is another length altogether, is believed, but not proved, to be order Z^0 as $Z \rightarrow \infty$.

The first correction to the $Z^{7/3}$ law is not, as was formerly supposed, the $Z^{5/3}$ corrections arising from exchange and correlation effects and kinetic energy corrections on the $Z^{-1/3}$ scale. Instead it is $Z^2 = Z^{6/3}$ and arises from extreme quantum mechanical effects on the innermost electrons, which are at a distance Z^{-1} from the nucleus. Among these the most important are the K shell electrons. It was Scott [19] who pointed this out and he also gave a formula for the correction term to the energy,

$$(1) \quad E^{\text{Scott}} = \frac{q}{8} Z^2,$$

where q is the number of spin states per electron (of course $q = 2$ in nature). It is noteworthy that E^{Scott} does not depend on the fixed ratio N/Z , provided $N/Z \neq 0$. This fact agrees with the idea that E^{Scott} arises from the innermost electrons, whose energies, in leading

¹⁾ Work partially supported by US National Science Foundation grant PHY90 19433 A04.

²⁾ Work partially supported by Norges forskningsråd, grant 412.92/008.

order, are independent of the presence of the electrons that are further from the nucleus. The truth of (1) (i.e., the statement that the energy is $E^{\text{TF}} + E^{\text{Scott}} + o(Z^2)$ for fixed $N/Z \neq 0$) was proved in [22], [23] (upper and lower bound) and by Hughes [9] (lower bound) in the atomic case and by Bach [1] in the ionic case. (For different proofs and extensions of this result see [24], Fefferman and Seco [6], [2], [3], [4], [5], [7], and Ivrii and Sigal [10].)

The conjecture about the density, which we prove here, was made later by Lieb in [11], Equation (5.34). It concerns the electron density ϱ_d at distances of order Z^{-1} from the nucleus and states that in limit $Z \rightarrow \infty$ a suitably scaled ϱ_d converges to the sum of the squares of *all* the hydrogenic bound states ϱ^H . Because of its relation to the Scott correction we will use henceforth the term “strong Scott correction”.

The function ϱ^H is defined in Section 2 below and is extensively analyzed in [8]. (Previously, an upper bound for ϱ_d at the origin of the correct form, namely $O(Z^3)$, was derived in [20], [21].) We prove this convergence in several senses, one of which is a “pointwise” convergence on spheres. In fact we go further and show that the individual angular momentum densities converge to the hydrogenic values, thereby giving a somewhat more refined picture of the ground state.

Our main proof strategy is the usual one. We add ε times a one-body test potential to our Hamiltonian and then differentiate the ground state energy with respect to ε at $\varepsilon = 0$ in order to find ϱ_d . To obtain pointwise convergence the test potential is a radial delta-function. To control the energy we rely, in part, on the results and methods in [22], [23].

In the following we shall state and prove our theorems for the neutral case $N = Z$. We do so to avoid the notational complexity and additional discussion required for $N/Z \neq 1$. It is straightforward, however, to generalize our results to $N/Z \neq 1$.

In the next section precise definitions, as well as our main theorems are given. Section 3 contains two lemmata about the difference in energies with and without the test potential. Since there are infinitely many hydrogenic bound states, we need these estimates in order to be able to show that the sum of the derivatives (with respect to ε) equals the derivative of the sum. The strong Scott conjecture for atoms is proved in Section 4. Section 5 contains the obvious extension to molecules. The Appendix contains a few needed facts about ground state energies.

2. Definitions and main results

The Hamiltonian of an atom of N electrons with q spin states each and a fixed nucleus of charge Z located at the origin is given by

$$(2) \quad H_{N,Z} = \sum_{v=1}^N \left(-\Delta_v - \frac{Z}{|\mathbf{r}_v|} \right) + \sum_{1 \leq \mu < v \leq N} \frac{1}{|\mathbf{r}_\mu - \mathbf{r}_v|},$$

in units in which $\hbar^2/2m = 1$ and $|e| = 1$. It is self-adjoint in the Hilbert space $\mathfrak{H}_N := \bigwedge_{v=1}^N (L^2(\mathbb{R}^3) \otimes \mathbb{C}^q)$, i.e., antisymmetric functions of space and spin. A general ground state density matrix, denoted by d , can be written as

$$(3) \quad d = \sum_{v=1}^M w_v |\psi_v\rangle \langle \psi_v|,$$

where the ψ_v constitute an orthonormal basis for the ground state eigenspace and where the w_v are nonnegative weights such that $\sum_{v=1}^M w_v = 1$. It is well known that the ground state can be degenerate, e.g., it is for the carbon atom. The corresponding one-electron density is the diagonal part of the one-electron density matrix and is, by definition,

$$(4) \quad \varrho_d(\mathbf{r}) = N \sum_{v=1}^M w_v \sum_{\sigma_1, \dots, \sigma_N=1}^q \int_{\mathbb{R}^{3(N-1)}} |\psi_v(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N; \sigma_1, \dots, \sigma_N)|^2 d\mathbf{r}_2 \dots d\mathbf{r}_N.$$

The density $\varrho_{l,d}$ of angular momentum l electrons at radius r from the nucleus is given in terms of the normalized spherical harmonics $Y_{l,m}(\omega)$.

$$(5) \quad \varrho_{l,d}(r) = N \sum_{m=-l}^l \sum_{v=1}^M w_v \sum_{\sigma_1, \dots, \sigma_N=1}^q \int_{\mathbb{R}^{3(N-1)}} |\int_{\mathbb{S}^2} \overline{Y_{lm}(\omega)} \psi_v(r\omega, \mathbf{r}_2, \dots, \mathbf{r}_N; \sigma_1, \dots, \sigma_N) d\omega|^2 d\mathbf{r}_2 \dots d\mathbf{r}_N$$

where we write $\mathbf{r} = r\omega$ and $d\omega$ denotes the usual unnormalized surface measure on the two dimensional sphere $\mathbb{S}^2$ with $(4\pi)^{-1} \int_{\mathbb{S}^2} d\omega = \int_{\mathbb{S}^2} |Y_{l,m}|^2 d\omega = 1$.

Throughout the paper we will write $\varphi_Z^{\text{TF}}(r)$ for the Thomas-Fermi potential of electron number $N = Z$ and nuclear charge Z , i.e.,

$$\varphi_Z^{\text{TF}}(r) = Z/r - |\mathbf{r}|^{-1} * \varrho_Z^{\text{TF}},$$

where ϱ_Z^{TF} is the nonnegative minimizer of the Thomas-Fermi functional

$$\int_{\mathbb{R}^3} \left(\frac{3}{5} (6\pi^2/q)^{2/3} \varrho^{5/3}(\mathbf{r}) - \frac{Z}{|\mathbf{r}|} \varrho(\mathbf{r}) \right) d\mathbf{r} + D(\varrho, \varrho)$$

under the condition $\int \varrho = N = Z$ and with

$$D(\varrho, \varrho) := \frac{1}{2} \int_{\mathbb{R}^6} \frac{\overline{\varrho(\mathbf{r})} \varrho(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} d\mathbf{r} d\mathbf{s}.$$

Both φ_Z and ϱ_Z^{TF} are spherically symmetric, i.e., they depend only on $r = |\mathbf{r}|$. There is a scaling relation $\varphi_Z^{\text{TF}}(r) = Z^{4/3} \varphi_1^{\text{TF}}(Z^{1/3} r)$, where φ_1^{TF} is the Thomas-Fermi potential for $Z = 1$ and “electron number” equal to 1. Similarly, $\varrho_Z^{\text{TF}}(r) = Z^2 \varrho_1^{\text{TF}}(Z^{1/3} r)$. This scaling shows that the “natural” length in an atom is $Z^{-1/3}$. Note that the Thomas-Fermi functional has a unique minimizer [12].

The Scott conjecture, on the other hand, concerns the length scale Z^{-1} , where we expect the density to be of order Z^3 instead of Z^2 . In terms of the “true” density defined in (4), we now define

$$(6) \quad \varrho_Z(\mathbf{r}) := Z^{-3} \varrho_d(\mathbf{r}/Z).$$

Likewise, we define the angular momentum density

$$(7) \quad \varrho_{l,Z}(r) := Z^{-3} \varrho_{l,d}(r/Z).$$

To formulate the strong Scott conjecture we consider the angular momentum l states of a hydrogen atom ($Z = 1$) with radial Hamiltonian

$$(8) \quad h_l^H := -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \frac{1}{r}$$

with normalized eigenfunctions $\psi_{l,n}^H$ (that vanish at 0 and ∞) corresponding to *negative* eigenvalues $e_{l,n}^H$. (The superscript H denotes “Hydrogen” and distinguishes h_l^H from other radial Hamiltonians to be considered later.) The normalization is $\int_0^\infty |\psi_{n,l}(r)|^2 dr = 1$. We define the corresponding density in the channel l to be

$$(9) \quad \varrho_l^H(r) := q(2l+1) \sum_{n=0}^{\infty} |\psi_{l,n}^H(r)|^2 / (4\pi r^2);$$

the total density is then

$$(10) \quad \varrho^H(r) = \sum_{l=0}^{\infty} \varrho_l^H(r).$$

Although we shall not be interested in detailed properties of ϱ^H , we note the following proved in [8]: The sum over l and n defining $\varrho^H(r)$ is pointwise convergent for all r . It is monotone decreasing and it decays asymptotically for large r as $1/(6\pi^2 r^{3/2})$. This large r asymptotics meshes nicely with the *small* r behavior of $\frac{1}{q} \varrho_1^{\text{TF}}(r)$.

Note: In [8] the operator h^H is defined using atomic units $\hbar^2/m = 1$, i.e., with $\frac{1}{2}(-d^2/dr^2 + l(l+1)/r^2)$ instead of $-d^2/dr^2 + l(l+1)/r^2$. Note also that we have included the factor q in the definition of ϱ which was not done in [8]. Thus some care is needed in comparing formulae there with formulae here.

Various notions for the convergence of the rescaled density are possible. Our precise statements are Theorems 1 and 2 below and Theorem 3 in Section 5.

Theorem 1 (Convergence in angular momentum channels). *Fix the angular momentum l_0 .*

1. *For positive r*

$$(11) \quad \lim_{Z \rightarrow \infty} \varrho_{l_0,Z}(r) = \varrho_{l_0}^H(r)$$

(pointwise convergence).

2. Let V be an integrable function on the positive real line. Then we have the weak convergence

$$(12) \quad \lim_{Z \rightarrow \infty} \int_0^\infty r V(r) q_{l_0, Z}(r) dr = \int_0^\infty r V(r) q_{l_0}^H(r) dr.$$

Theorem 2 (Convergence of the total density). 1. Let W be a bounded (not necessarily constant) function on the unit sphere and r positive. Then, as $Z \rightarrow \infty$,

$$(13) \quad \int_{\mathbb{S}^2} W(\omega) q_Z(r\omega) d\omega \rightarrow q^H(r) \int_{\mathbb{S}^2} W(\omega) d\omega$$

(pointwise convergence of spherical averages).

2. Let V be a locally bounded, integrable function on $\mathbb{R}^3$. Then, as $Z \rightarrow \infty$,

$$(14) \quad \int_{\mathbb{R}^3} |\mathbf{r}| V(\mathbf{r}) q_Z(\mathbf{r}) d\mathbf{r} \rightarrow \int_{\mathbb{R}^3} |\mathbf{r}| V(\mathbf{r}) q^H(|\mathbf{r}|) d\mathbf{r}.$$

Remarks. 1. It is not really necessary to take a sequence of ground state density matrices. We could take just a sequence of states, $d_{N,Z}$, that is an *approximate ground state* in the sense that

$$\frac{\text{tr}(H_{N,Z} d_{N,Z}) - E_{N,Z}}{Z^2} \rightarrow 0$$

as $Z \rightarrow \infty$. Here $E_{N,Z}$ is the bottom of the spectrum of $H_{N,Z}$. It might not be an eigenvalue, and it certainly will not be one if N/Z is larger than 2.

2. It is important to note that W and V in (13) and (14) need not be spherically symmetric. It might appear that only the spherical averages of W and V are relevant, but this would miss the point. Theorem 2 says, that in the limit $Z \rightarrow \infty$ there is no way to construct a ground state or approximate ground state that is not spherically symmetric on a length scale Z^{-1} . For example, in the case of carbon there are ground states that are not spherically symmetric and for which replacing W by its spherical average changes the left side of (13).

3. A word about pointwise convergence. The one-body density matrix $\gamma(\mathbf{r}, \mathbf{r}')$ which is defined as in (4) but with $|\psi_v(\cdots)|^2$ replaced by

$$\psi_v(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N; \sigma_1, \dots, \sigma_N) \overline{\psi_v(\mathbf{r}', \mathbf{r}_2, \dots, \mathbf{r}_N; \sigma_1, \dots, \sigma_N)},$$

is easily seen to be in the Sobolev space $H^1(\mathbb{R}^3)$ when γ is considered as a function of each variable separately. The trace theorem in Sobolev spaces then implies that the function of ω on the sphere $\mathbb{S}^2$, $\gamma(r\omega, r'\omega')$ is in $L^q(\mathbb{S}^2)$ for all $q \leq 4$. Thus, the integrals in (4) are well defined and $q_d(r\omega) = \gamma(r\omega, r\omega)$ is in $L^2(\mathbb{S}^2)$. It is also easy to see that $\sqrt{q_{l,Z}}(r)$ is in $H^1(0, \infty)$ and hence it is a continuous function of r . Since $q_d(\mathbf{r})$ is in $L^2(\mathbb{S}^2)$ the integrals in (11) and (13) are well defined when $W \in L^2(\mathbb{S}^2)$.

If q and γ belong to a ground state of $H_{N,Z}$ with $N \leq Z$ then they are even continuous functions in all variables. This follows from the regularity theorem of Kato and Simon

(Reed and Simon [18], Theorem XIII.39) and the uniform exponential decay of ground state eigenfunctions. This decay is implied as follows. By Zhislin's theorem the atomic Hamiltonian has infinitely many eigenvalues below the essential spectrum and the ground state eigenspace has finite dimension. This implies that the ground state energy is always a discrete eigenvalue which, in turn, implies exponential decay of the ground state eigenfunctions according to Theorem XIII.42 of [18].

3. Eigenvalue differences of Schrödinger operators perturbed on the scale $1/Z$

It is well known, and will be seen more explicitly in Section 4, that the eigenvalues of $H_{N,Z}$ can be controlled to within an accuracy of $o(Z^2)$ by considering a one-body Schrödinger operator with the spherically symmetric potential given by Thomas-Fermi theory. In the angular momentum l channel, this is

$$(15) \quad h_{l,Z}^{\text{TF}} = -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \varphi_Z^{\text{TF}}(r).$$

(We suppress the dependence on N in $h_{l,Z}^{\text{TF}}$ since $N = Z$.) Closely related to $h_{l,Z}^{\text{TF}}$ is the unscreened hydrogenic Hamiltonian

$$(16) \quad h_{l,Z}^{\text{H}} = -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \frac{Z}{r}.$$

In this section we want to study how the spectra of these operators are shifted by the addition of a perturbing potential of the form

$$\varepsilon U_Z(r) = \varepsilon Z^2 U(Zr)$$

where ε is a small parameter and where U is some fixed function. In particular, U will be a radial delta function, $U(r) = \delta(r-a)$ for some $a > 0$.

Both cases, h^{TF} and h^{H} , will be considered together and we write

$$(17) \quad h_{l,\varepsilon,Z} = -\frac{d^2}{dr^2} + V_{l,Z}(r) - \varepsilon U_Z(r),$$

in which $V_{l,Z} = -Z/r + l(l+1)/r^2$ or $V_{l,Z} = -\varphi_Z^{\text{TF}}(r) + l(l+1)/r^2$.

Our first lemma estimates the difference in the spectra of $h_{l,0,Z}$ and $h_{l,\varepsilon,Z}$ by the difference in the trace (tr) of the negative parts $(h_{l,\varepsilon,Z})_-$ and $(h_{l,0,Z})_-$ (i.e., the sums of the negative eigenvalues). This lemma will later on allow us to interchange the limits $Z \rightarrow \infty$ and $\varepsilon \rightarrow 0$ with the l summation.

Lemma 1. *Set $U(r) = \delta(r-a)$, $U_Z(r) = Z^2 U(Zr)$ and assume $|\varepsilon| \leq \pi/(16a)$. Then*

$$|\text{tr}(h_{l,0,Z})_- - \text{tr}(h_{l,\varepsilon,Z})_-| \leq |\varepsilon| \frac{9aZ^2}{(l+1)^2(2l+1)}.$$

Proof. By the minimax principle we have for $\varepsilon > 0$

$$(18) \quad 0 \leq s_{\varepsilon,l,Z} := \operatorname{tr}(h_{l,0,Z}) - \operatorname{tr}(h_{l,\varepsilon,Z}) - \leq \varepsilon \operatorname{tr}(U_Z d_{l,\varepsilon,Z}).$$

Inserting the identity twice in the right side of (10) we have

$$(19) \quad s_{\varepsilon,l,Z} \leq \varepsilon \operatorname{tr}(A \circ B \circ C \circ B^* \circ A^*) \leq \varepsilon \|A\|_\infty^2 \|B\|_\infty^2 \operatorname{tr} C$$

with

$$\begin{aligned} A &:= d_{l,\varepsilon,Z}(h_{l,\varepsilon,Z} + c_{l,Z})^{1/2} \geq 0, \\ B &:= (h_{l,\varepsilon,Z} + c_{l,Z})^{-1/2}(H_{0,l} + c_{l,Z})^{1/2}, \\ C &:= (H_{0,l} + c_{l,Z})^{-1/2} U_Z (H_{0,l} + c_{l,Z})^{-1/2} \geq 0, \end{aligned}$$

where $c_{l,Z}$ is any positive number bigger than $|\inf \sigma(h_{l,\varepsilon,Z})|$, where $\sigma(h)$ denotes the spectrum of h . We also define $H_{0,l} := -d^2/dr^2 + l(l+1)/r^2$ to be the free operator in the angular momentum channel l . Since $\varphi_Z^{\text{TF}}(r) \leq Z/r$ and since $\inf \sigma(H_{0,l} - Z/r) = -Z^2/[4(l+1)^2]$ we can take $c_{l,Z} := 2Z^2/(l+1)^2$ provided ε is not too large.

We now estimate these norms individually:

Because $c_{l,Z}$ is bigger than the modulus of the lowest spectral point of $h_{l,\varepsilon,Z}$ and $d_{l,\varepsilon,Z}$ is the projection onto the negative spectral subspace of $h_{l,\varepsilon,Z}$ we have

$$(20) \quad \|A\|_\infty \leq \sqrt{c_{l,Z}}.$$

For B we get

$$\begin{aligned} (21) \quad \|B\phi\|^2 &= \|(h_{l,\varepsilon,Z} + c_{l,Z})^{-1/2}(H_{0,l} + c_{l,Z})^{1/2}\phi\|^2 \\ &= (\phi, (H_{0,l} + c_{l,Z})^{1/2}(h_{l,\varepsilon,Z} + c_{l,Z})^{-1}(H_{0,l} + c_{l,Z})^{1/2}\phi) \\ &= \left(\phi, \frac{1}{1 - W_{l,\varepsilon,Z}} \phi \right) \end{aligned}$$

with

$$W_{l,\varepsilon,Z} := (H_{0,l} + c_{l,Z})^{-1/2}(\varphi_Z + \varepsilon U_Z)(H_{0,l} + c_{l,Z})^{-1/2}.$$

We will then have

$$(22) \quad \|B\| \leq \sqrt{2}$$

if we can show that $W_{l,\varepsilon,Z}$ is bounded above by $\frac{1}{2}$. To this end we note that $H_{0,l} + c_{l,Z}$ is invertible, so that we can write any normalized $\phi \in L^2(\mathbb{R}^+)$ as

$$\phi := (H_{0,l} + c_{l,Z})^{1/2} \psi / \|(H_{0,l} + c_{l,Z})^{1/2} \psi\|$$

with ψ in the domain of $H_{0,l}$. Thus, we have to show that

$$(\phi, W_{l,\varepsilon,Z} \phi) = (\psi, (\varphi_Z + \varepsilon U_Z) \psi) / (\psi, (H_{0,l} + c_{l,Z}) \psi) \leq \frac{1}{2},$$

which is equivalent to

$$(23) \quad \frac{1}{2} (H_{0,l} + c_{l,Z}) - \varphi_Z - \varepsilon U_Z \geq 0.$$

Since $\varphi_Z^{\text{TF}}(r) \leq Z/r$ and since

$$(24) \quad (\psi, U_Z \psi) = Z |\psi(a/Z)|^2 \leq Z \int_0^{a/Z} |\psi|^2(r) dr \leq 2Z \Re \int_0^{a/Z} \psi(r) \psi'(r) dr \\ \leq 2Z \|\psi'\|_2 \left\{ \int_0^{a/Z} \psi(r)^2 dr \right\}^{1/2}$$

we have that $(\psi, U_Z \psi) \leq (4a/\pi) \|\psi'\|_2^2$. (Here we use the inequality, $\int_0^L \psi'^2 \geq (\pi/2L)^2 \int_0^L \psi^2$, when $\psi(0) = 0$.) Thus, (23) is implied by

$$(25) \quad -4 \frac{Z^2}{4(l+1)^2} + \frac{c_{l,Z}}{2} + \inf \sigma \left(\left(\frac{1}{4} - \frac{4\varepsilon a}{\pi} \right) H_{0,l} \right) \geq 0.$$

The sum of the first two terms in (25) vanishes because of our choice of $c_{l,Z}$ and last term in (25) is zero when $\varepsilon \leq \pi/(16a)$. This is true by hypothesis, and the bound on the norm of B is proved.

Finally the trace of C is computed easily, since it is of rank one. Since the kernel $(H_{0,l} + c_{l,Z})^{-1}(r, r')$, is a positive, continuous function in both variables,

$$\text{tr } C = Z(H_0 + c_{l,Z})^{-1}(a/Z, a/Z).$$

A well known calculation yields

$$(H_{0,l} + c_{l,Z})^{-1}(r, r') = \sqrt{r} K_{l+\frac{1}{2}}(\sqrt{c_{l,Z}} r_>) I_{l+\frac{1}{2}}(\sqrt{c_{l,Z}} r_<) \sqrt{r'},$$

where $r_> = \max\{r, r'\}$ and $r_< = \min\{r, r'\}$. Thus

$$\text{tr } C = a K_{l+\frac{1}{2}} \left(\sqrt{c_{l,Z}} \frac{a}{Z} \right) I_{l+\frac{1}{2}} \left(\sqrt{c_{l,Z}} \frac{a}{Z} \right).$$

The modified Bessel functions $I_{l+\frac{1}{2}}$ and $K_{l+\frac{1}{2}}$ are both positive and the following uniform asymptotic expansions hold. (See Olver [14] for a proof of the estimates of the remainder terms, [15], p. 6 for the remainder in the form used here, [17], Chapter 10, Paragraph 7 for a review; see also Olver [16], section 9.7.)

$$(26) \quad K_n(nx) = \sqrt{\frac{\pi t}{2n}} e^{-n\xi} [1 + \varepsilon_{0,2}(n, t)],$$

$$(27) \quad I_n(nx) = \sqrt{\frac{t}{2\pi n}} \frac{e^{n\xi}}{1 - \varepsilon_{0,1}(n, 0)} [1 + \varepsilon_{0,1}(n, t)],$$

where

$$\xi := \frac{1}{t} - \frac{1}{2} \log \frac{1+t}{1-t},$$

$$t := (1 + x^2)^{-\frac{1}{2}}$$

and

$$|\varepsilon_{0,1}(n, t)| \leq \frac{n_0}{n - n_0},$$

$$|\varepsilon_{0,2}(n, t)| \leq \frac{n_0}{n - n_0}$$

with $n_0 := \frac{1}{6\sqrt{5}} + \frac{1}{12} \leq \frac{1}{6}$. Thus

$$K_n(nx) I_n(nx) \leq \frac{1}{2n(1+x^2)^{\frac{1}{2}}} \frac{n^2}{(n-2n_0)(n-n_0)} \leq \frac{9}{4n},$$

where the last inequality holds for $n \geq \frac{1}{2}$. Thus

$$(28) \quad \text{tr } C \leq \frac{9}{2} \frac{a}{(2l+1)}.$$

Putting (19), (20), (22), and (28) together yields

$$s_{\varepsilon,l,z} \leq \varepsilon \frac{9c_{l,z}a}{(2l+1)} \leq \varepsilon \frac{9aZ^2}{(l+1)^2(2l+1)}$$

which is more than the desired result for $\varepsilon > 0$.

If ε is negative we have, again by the minimax principle,

$$(29) \quad 0 \geq s_{\varepsilon,l,z} := \text{tr}(h_{l,0,z}) - \text{tr}(h_{l,\varepsilon,z}) - \geq \varepsilon \text{tr}(U_Z d_{l,0,z}).$$

Similar to the previous analysis, we have

$$(30) \quad s_{\varepsilon,l,z} \geq \varepsilon \text{tr}(D \circ E \circ C \circ E^* \circ D^*) \geq \varepsilon \|D\|_{\infty}^2 \|E\|_{\infty}^2 \text{tr } C$$

with $D := d_{l,0,z}(h_{l,0,z} + c_{l,z})^{\frac{1}{2}}$ and $E = (h_{l,0,z} + c_{l,z})^{-\frac{1}{2}}(H_{l,0} + c_{l,z})^{\frac{1}{2}}$. As for A above, we have $\|D\|_{\infty} \leq \sqrt{c_{l,z}}$ and $\|E\| \leq \sqrt{2}$. Putting this together with (28) gives

$$s_{\varepsilon,l,z} \geq \varepsilon \frac{9c_{l,z}a}{(2l+1)} \geq \varepsilon \frac{9aZ^2}{(l+1)^2(2l+1)}$$

which is the desired result for negative ε . $\square$

The next result will later on allow us to interchange the limits $Z \rightarrow \infty$ and $\varepsilon \rightarrow 0$ with the n summation for fixed l .

Lemma 2. Set $U(r) = \delta(r-a)$ and assume $|\varepsilon| \leq \pi/(4a)$, $a > 0$. Let

$$(31) \quad h_{l,\varepsilon} := -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \frac{1}{r} - \varepsilon U(r)$$

with form domain $H_0^1(0, \infty)$. Let $e_{n,l,\varepsilon}$ denote the n -th eigenvalue of $h_{l,\varepsilon}$. Then

$$(32) \quad |e_{n,l,0} - e_{n,l,\varepsilon}| \leq \frac{1}{(n+l)^2} \frac{|\varepsilon|a}{\pi - 4\varepsilon a}.$$

Proof. For any ψ in $H_0^1(0, \infty)$ we have

$$|\psi(a)|^2 \leq 2 \frac{2}{\pi} a \|\psi'\|_2^2,$$

as proved in (24) of Lemma 1. Thus, for $\varepsilon > 0$,

$$(33) \quad h_{l,\varepsilon} \geq \left(1 - \frac{4\varepsilon a}{\pi}\right) \left[-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \frac{1}{\left(1 - \frac{4\varepsilon a}{\pi}\right)r}\right].$$

This implies

$$e_{n,l,\varepsilon} \geq \left(1 - \frac{4\varepsilon a}{\pi}\right) \tilde{e}_{n,l,0}$$

where $\tilde{e}_{n,l,0}$ is the n -th eigenvalue of $[\]$ in (33), i.e., where the potential r^{-1} is replaced by $(1 - 4\varepsilon a/\pi)^{-1}r^{-1}$. Thus,

$$0 \leq e_{n,l,0} - e_{n,l,\varepsilon} \leq \frac{1}{4(n+l)^2} \left(-1 + \left(1 - \frac{4\varepsilon a}{\pi}\right)^{-1}\right) = \frac{1}{(n+l)^2} \frac{\varepsilon a}{\pi - 4\varepsilon a},$$

which proves the claim when $0 < \varepsilon < \pi/(4a)$.

If ε is negative we have

$$h_{l,\varepsilon} \leq \left(1 - \frac{4\varepsilon a}{\pi}\right) \left[-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \frac{1}{\left(1 - \frac{4\varepsilon a}{\pi}\right)r}\right]$$

which again proves the claim (by the same argument) when $0 > \varepsilon > -\pi/(4a)$. $\square$

4. Proof of the strong Scott conjecture

We are now able to give the proofs of our theorems. We begin with the proof of Theorem 1 and begin with the first statement:

1. *The proof of the convergence of the spherical averages.* Set $U(r) := \delta(r - a)$ for $a > 0$ and $U_Z(r) := Z^2 U(Zr) = Z\delta\left(r - \frac{a}{Z}\right)$. Fix l_0 and let

$$(34) \quad H_{N,Z}^\varepsilon := H_{N,Z} - \varepsilon \sum_{v=1}^N U_Z(r_v) \Pi(l_0)$$

where $\Pi(l_0)$ denotes the projection onto angular momentum l_0 . We define $\lambda(Z)$ – which does not depend on ε – by

$$(35) \quad \lambda(Z) := a^2 \varrho_{l_0}(a) = \frac{\text{tr}(H_{N,Z} d) - \text{tr}(H_{N,Z}^\varepsilon d)}{\varepsilon Z^2}.$$

Let us define $e_{n,l,\gamma,Z}^H$ and $e_{n,l,\gamma,Z}$, $n = 1, 2, \dots, \varepsilon \in \mathbb{R}$, to be the negative eigenvalues of the operators

$$(36) \quad H_{l,\varepsilon,Z}^H := -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \frac{Z}{r} - \varepsilon U_Z \delta_{l,l_0},$$

$$(37) \quad H_{l,\varepsilon,Z} := -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} - \varphi_Z^{\text{TF}} - \varepsilon U_Z \delta_{l,l_0}$$

with zero Dirichlet boundary on $(0, \infty)$. To obtain an upper bound for $\lambda(Z)$ we pick ε positive and estimate as follows: by (63) we have the upper bound

$$(38) \quad \begin{aligned} & \text{tr}(H_{N,Z} d) \\ & \leq \sum_{l=0}^{L-1} q(2l+1) \sum_n e_{n,l,0,Z}^H + \sum_{l=L}^{\infty} q(2l+1) \sum_n e_{n,l,0,Z} - D(\varrho_{\text{TF}}, \varrho_{\text{TF}}) + \text{const } Z^{\frac{47}{24}} \end{aligned}$$

where $L = \lceil Z^{1/9} \rceil$.

To obtain a lower bound on $\text{tr}(H_{N,Z}^\varepsilon d)$ we first use the lower bound ([13], [11]) on the correlations, namely $-\text{const} [N \int \varrho_t^{5/3}]^{1/2}$, to reduce it to a radial problem. Using the fact that $Z/r \geq \varphi_Z^{\text{TF}}(r)$ for $r > 0$ it follows from this that

$$(39) \quad \begin{aligned} & \text{tr}(H_{N,Z}^\varepsilon d) \\ & \geq \sum_{l=0}^{L-1} q(2l+1) \sum_n e_{n,l,\varepsilon,Z}^H + \sum_{l=L}^{\infty} q(2l+1) \sum_n e_{n,l,\varepsilon,Z} - D(\varrho_{\text{TF}}, \varrho_{\text{TF}}) - \text{const } Z^{\frac{5}{3}}. \end{aligned}$$

Note that (39) arises from a relatively simple lower bound calculation. Part of the proof of the Scott conjecture amounts to proving that the right hand of (39) is accurate to $o(Z^2)$. This proof was carried out in [23] (see also [9], [24]). We are *not* rederiving the Scott correction for the energy, and it is not necessary for us to do so here.

Define

$$\theta(n) := \begin{cases} 1, & n > 0, \\ 0, & n \leq 0. \end{cases}$$

Since the eigenvalues of the perturbed problem ($\varepsilon \neq 0$) are equal to the unperturbed one ($\varepsilon = 0$) except for $l = l_0$, we get the inequality

$$\begin{aligned} (40) \quad & \limsup_{Z \rightarrow \infty} \lambda(Z) \\ & \leq \liminf_{\varepsilon \searrow 0} \limsup_{Z \rightarrow \infty} \left[q(2l_0 + 1) \frac{\operatorname{tr}(H_{l_0,0,Z}^H) - \operatorname{tr}(H_{l_0,\varepsilon,Z}^H)}{\varepsilon Z^2} \theta(L - l_0) \right. \\ & \quad \left. + \frac{\operatorname{tr}(H_{l_0,0,Z}) - \operatorname{tr}(H_{l_0,\varepsilon,Z})}{\varepsilon Z^2} \theta(l_0 - L) + \operatorname{const} Z^{-\frac{1}{24}} \varepsilon^{-1} \right]. \end{aligned}$$

Because L eventually becomes larger than the fixed l_0 ,

$$\begin{aligned} (41) \quad & \limsup_{Z \rightarrow \infty} \lambda(Z) \leq q(2l_0 + 1) \liminf_{\varepsilon \searrow 0} \limsup_{Z \rightarrow \infty} \frac{\operatorname{tr}(H_{l_0,0,Z}^H) - \operatorname{tr}(H_{l_0,\varepsilon,Z}^H)}{\varepsilon Z^2} \\ & = q(2l_0 + 1) \liminf_{\varepsilon \searrow 0} \frac{\operatorname{tr}(H_{l_0,0,1}^H) - \operatorname{tr}(H_{l_0,\varepsilon,1}^H)}{\varepsilon}, \end{aligned}$$

where the last equation holds since because of the scaling of $h_{l_0,0,Z}^H$ and $h_{l_0,\varepsilon,Z}$. Therefore

$$(42) \quad \limsup_{Z \rightarrow \infty} \lambda(Z) \leq q(2l_0 + 1) \liminf_{\varepsilon \searrow 0} \sum_n \frac{e_{n,l_0,0,1}^H - e_{n,l_0,\varepsilon,1}^H}{\varepsilon}$$

$$(43) \quad = q(2l_0 + 1) \sum_n \liminf_{\varepsilon \searrow 0} \frac{e_{n,l_0,0,1}^H - e_{n,l_0,\varepsilon,1}^H}{\varepsilon}$$

$$(44) \quad = a^2 \varrho_{l_0}^H(a).$$

To exchange the limit $\varepsilon \searrow 0$ with the summation in (42) we use Lemma 2, which provides a summable majorant for the series that is uniform in ε and thus allows us to fulfil Weierstraß' criterion for uniform convergence. Finally, to deduce (44) from (43) we use the fact that the one-dimensional delta potential is a relatively form bounded perturbation, i.e., defines an analytic family in the sense of Kato.

To obtain a lower bound for $\lambda(Z)$ we pick ε negative instead of positive and take the limit $\limsup_{\varepsilon \nearrow 0}$ and $\liminf_{Z \rightarrow \infty}$ instead of $\liminf_{\varepsilon \searrow 0}$ and $\limsup_{Z \rightarrow \infty}$. Repeating the same steps gives the same result except for reversing the inequalities, thereby yielding the same bound (44) from below. This establishes the first claim of the theorem.

2. Proof of the weak convergence. Because of the linear dependence of the right and left hand side of (12), it suffices to prove the claim for the positive and negative parts of V separately. Thus we may – and shall – assume that V is positive. We can now roll the

proof back to the previous case as follows: First we pick Z large enough so that $l_0 < L$. It is convenient now, to replace ε by ε/a in order that the right side of (32) in Lemma 2 is uniformly bounded in a and ε for all $a \in (0, \infty)$ and for $|\varepsilon| \leq \pi/8$. Then we integrate the inequality

$$(45) \quad a^2 q_{l_0}(a) \leq \frac{\operatorname{tr}(H_{l_0,0,Z}^H) - \operatorname{tr}(H_{l_0,\varepsilon/a,Z}^H)}{(\varepsilon/a)Z^2} \theta(L-l) + \operatorname{const}/((\varepsilon/a)Z^{\frac{1}{24}}),$$

i.e.,

$$(46) \quad a q_{l_0}(a) \leq \frac{\operatorname{tr}(H_{l_0,0,Z}^H) - \operatorname{tr}(H_{l_0,\varepsilon/a,Z}^H)}{\varepsilon Z^2} \theta(L-l) + \operatorname{const}/(\varepsilon Z^{\frac{1}{24}})$$

against $V(a)$ from 0 to ∞ . Thanks to Lemma 1 the right side of (46) is bounded by $\operatorname{const} a$ and hence the integral is finite. Next we write out the traces appearing in (44) in terms of the eigenvalues and then use Lemma 2 to provide a bound that is summable (over the eigenvalues) and integrable (from 0 to ∞), if $|\varepsilon| < \pi/8$. This bound is uniformly bounded in ε , and so, by dominated convergence, we can take the limit $\varepsilon \searrow 0$ term by term. Using the result (11), which we established above, Equation (12) is now verified. $\square$

Proof of Theorem 2. As was the case in the proof of Theorem 2 we shall assume that W and V are nonnegative. For Part 1 we proceed as for Theorem 2 and define $H_{N,Z}^\varepsilon$ as in (34), but with $\Pi(l_0)$ replaced by $W(\omega)$. First we treat the case $W(\omega) = 1$. We follow the proof of Theorem 1 up to equations (36) and (37) (with δ_{l,l_0} replaced by W). Then we obtain analogously

$$(47) \quad \limsup_{Z \rightarrow \infty} \lambda(Z)$$

$$(48) \quad \leq \liminf_{\varepsilon \searrow 0} \limsup_{Z \rightarrow \infty} \left[\sum_{l=0}^{\infty} q(2l+1) \frac{\operatorname{tr}(H_{l,0,Z}^H) - \operatorname{tr}(H_{l,\varepsilon,Z}^H)}{\varepsilon Z^2} \theta(L-l) \right]$$

$$(49) \quad + \sum_{l=0}^{\infty} q(2l+1) \frac{\operatorname{tr}(H_{l,0,Z}) - \operatorname{tr}(H_{l,\varepsilon,Z})}{\varepsilon Z^2} \theta(l-L) + \operatorname{const} Z^{-\frac{1}{24}} \varepsilon^{-1}$$

$$(50) \quad = \sum_{l=0}^{\infty} q(2l+1) \liminf_{\varepsilon \searrow 0} \limsup_{Z \rightarrow \infty} \frac{\operatorname{tr}(H_{l,0,Z}^H) - \operatorname{tr}(H_{l,\varepsilon,Z}^H)}{\varepsilon Z^2}$$

$$(51) \quad = \sum_{l=0}^{\infty} q(2l+1) \liminf_{\varepsilon \searrow 0} \frac{\operatorname{tr}(H_{l,0,1}^H) - \operatorname{tr}(H_{l,\varepsilon,1}^H)}{\varepsilon}$$

$$(52) \quad = a^2 \sum_{l=0}^{\infty} \liminf_{\varepsilon \searrow 0} q_l^H(a)$$

$$(53) \quad = a^2 q^H(a).$$

To obtain (49) we use inequalities (38) and (39). To obtain (50) we use the fact that Lemma 1 provides a majorant uniform in ε and Z that is absolutely summable with respect to $\sum_{l=0}^{\infty} q(2l+1)$, i.e., fulfills the Weierstraß criterion for uniform convergence (or the hypothesis of Lebesgue's dominated convergence theorem), and therefore allows the interchange of

the limit and the l summation, and that the second sum tends term by term to zero. To obtain (51) we use the fact that the eigenvalues of the bare problem scale like Z^2 . Finally, the convergence result of Theorem 2 was used to obtain (52).

To obtain a lower bound we pick ε negative instead of positive and take the limit $\limsup_{\varepsilon \nearrow 0}$ and $\liminf_{Z \rightarrow \infty}$ instead of $\liminf_{\varepsilon \searrow 0}$ and $\limsup_{Z \rightarrow \infty}$. Repeating the same steps gives the same result except for reversing the inequalities thereby yielding the bound from below.

Let W now be a general bounded, measurable function on the unit sphere which we may – according to the remarks in the beginning – assume to be positive. We take $\|W\|_\infty = 1$.

Let us try to imitate the steps (47) to (53). As before we are faced with estimating the eigenvalues of the one-body operators

$$(54) \quad H_{\varepsilon,Z}^H := -\Delta - Z/|\cdot| - \varepsilon ZW\delta_Z^a \quad \text{and} \quad H_{\varepsilon,Z}^{\text{TF}} = -\Delta - \varphi_Z^{\text{TF}} - \varepsilon ZW\delta_Z^a$$

but unlike the previous case they cannot be simply indexed by the angular momentum l when $\varepsilon \neq 0$; indeed the one-body operators cannot be reduced to a direct sum of radial Schrödinger operators as in (36) and (37). However, the eigenvalues are real analytic functions of ε and we can label the eigenvalues by the l -value they have when ε tends to zero. In short, the only change needed in (47) to (53) is to replace $(2l+1)e_{n,l,\varepsilon,Z}^H$ by the sum of the eigenvalues in the multiplet of $H_{\varepsilon,Z}^H$ that converge to $e_{n,l,0,Z}^H$ as ε tends to zero. Since W is bounded by 1, all our previous bounds for eigenvalue differences (Lemmata 1 and 2) continue to hold and we are finally led to the $\liminf_{\varepsilon \searrow 0}$ in (51).

The crucial point is this: Even if W is not spherical symmetric, the sum of the eigenvalues in any multiplet is rotationally invariant to first order in ε in the following sense. The only property of W that matters – to first order – is the average $W_{\text{average}} := (4\pi)^{-1} \int W(\omega) d\omega$.

Reversing the sign of ε again gives the lower bound.

2. Proof of the weak convergence. The proof can be rolled back to the previous case as follows: First we assume that V is spherically symmetric and integrate the inequality

$$\begin{aligned} a^2 \varrho_Z(a\omega) \leq & \left[\sum_{l=0}^{\infty} q(2l+1) \frac{\text{tr}(H_{l,0,Z}^H) - \text{tr}(H_{l,\varepsilon,Z}^H)}{\varepsilon Z^2} \theta(L-l) \right. \\ & \left. + \sum_{l=0}^{\infty} q(2l+1) \frac{\text{tr}(H_{l,0,Z}) - \text{tr}(H_{l,\varepsilon,Z})}{\varepsilon Z^2} \theta(l-L) \right] + \text{const}/(Z^{\frac{1}{4}}) \end{aligned}$$

against $aV(a)$ from 0 to ∞ . Observe that because of Lemma 1 the summand of the sum on the right side of this integrated inequality is uniformly bounded by

$$\frac{9}{(l+1)^2(2l+1)} \int_0^\infty V(a) a^2 da$$

which, when multiplied by $(2l + 1)$, is summable. Again, the same argument holds when expressing the traces as sum over eigenvalues. Thus we are allowed to take the limits term by term for the differences of the eigenvalues giving the desired result as above.

The extension to the non-spherical case is as in Part 1. $\square$

5. Extensions to molecules

The ground state energy of a neutral molecule with nuclear charges $Z_1 = \lambda z_1, \dots, Z_K = \lambda z_K$ and positions of the nuclei at $\mathfrak{R}_1, \dots, \mathfrak{R}_K$ is given as

$$(55) \quad E(N, \vec{Z}) = \inf \{ \inf \sigma(H_{N, \vec{Z}, \vec{R}}) \mid \vec{R} \in \mathbb{R}^{3K} \}$$

where

$$(56) \quad H_{N, \vec{Z}, \vec{R}} = \sum_{v=1}^N \left(-\Delta_v - \sum_{\kappa=1}^K \frac{Z_\kappa}{|\mathbf{r}_v - \mathfrak{R}_\kappa|} \right) + \sum_{\substack{\mu, v=1 \\ \mu < v}}^N \frac{1}{|\mathbf{r}_\mu - \mathbf{r}_v|} + \sum_{\substack{\kappa, \kappa'=1 \\ \kappa < \kappa'}} \frac{Z_\kappa Z_{\kappa'}}{|\mathfrak{R}_\kappa - \mathfrak{R}_{\kappa'}|}$$

self-adjointly realized in $\mathfrak{H}_N$. Here $\vec{Z}$ denotes the K -tuple $(Z_1, \dots, Z_K)$ and $\vec{R}$ the $3K$ -tuple $(\mathfrak{R}_1, \dots, \mathfrak{R}_K)$. We also set $\vec{z} := (z_1, \dots, z_K)$. Solovej [25] showed recently that for arbitrary but fixed $\vec{z}$ and $N = Z_1 + \dots + Z_K$

$$(57) \quad E(N, \vec{Z}) = \sum_{\kappa=1}^K E(Z_\kappa, z_\kappa) + o(\lambda^{\frac{5}{3}})$$

holds as λ tends to infinity and that the minimizing inter-nuclear distances are of order $\lambda^{-5/21}$ or bigger. These results imply among other things not only that the atomic Scott correction and Schwinger correction implies the molecular one but allows us to generalize Theorem 2 as well: The molecular density in the vicinity of each nucleus converges in the sense of Theorem 1 to the hydrogen density at each of the centers. Our precise result is:

Theorem 3. Assume that $E(N, \vec{Z})$ as defined in (55) is equal to

$$(58) \quad \inf \{ \inf \sigma(H_{N, \vec{Z}, \vec{R}}) \mid \vec{R} \in \mathbb{R}^{3K}, \forall 1 \leq \kappa < \kappa' \leq K, |\mathfrak{R}_\kappa - \mathfrak{R}_{\kappa'}| \geq R := \text{const } \lambda^\gamma \}$$

with $\gamma > -1/4$. Assume $N = Z_1 + \dots + Z_K$, $Z_1 = \lambda z_1, \dots, Z_K = \lambda z_K$ with given fixed $z_1, \dots, z_K$. Furthermore fix $\kappa_0 \in 1, \dots, K$ and pick a sequence of ground state density matrices d_c of $H_{N, \vec{Z}, \vec{R}}$ with densities q_{d_c} . Define $q_{\lambda, \kappa_0}(r) := q_\lambda((r - \mathfrak{R}_{\kappa_0})/\lambda)/\lambda^3$. Finally assume $W \in L^2(\mathbb{S}^2)$. Then for $r > 0$

$$(59) \quad \int_{\mathbb{S}^2} W(\omega) q_{\lambda, \kappa_0}(r\omega) \rightarrow q q^H(r) \int_{\mathbb{S}^2} W$$

as $\lambda \rightarrow \infty$.

Proof. First note that by suitable relabeling we can always assume that $\kappa_0 = 1$. Set

$$H_{N, \vec{Z}, \vec{R}}^e := H_{N, \vec{Z}, \vec{R}} - \sum_{v=1}^N \varepsilon U_\lambda(\mathbf{r}_v) \text{ where } U_\lambda \text{ is defined as in the atomic case, i.e.,}$$

$$U_\lambda(\mathbf{r}) := \lambda^2 U(\lambda(\mathbf{r} - \mathfrak{R}_1)), \quad U(\mathbf{r}) := W(\omega) \delta(r - a),$$

and W is a square integrable function on the unit sphere. Because of (57) it suffices that

$$\mathrm{tr}(dH_{N,\bar{Z},\bar{R}}^\varepsilon) \geq \mathrm{tr}(H_{\varepsilon,Z_1}^{\mathrm{TF}} - D(\varrho_{Z_1}^{\mathrm{TF}}, \varrho_{Z_1}^{\mathrm{TF}})) + \sum_{\kappa=2}^N (\mathrm{tr}(H_{0,Z_\kappa}^{\mathrm{TF}} - D(\varrho_{Z_\kappa}^{\mathrm{TF}}, \varrho_{Z_\kappa}^{\mathrm{TF}})) - \mathrm{const} Z^{2-\delta})$$

for some positive δ and an approximate ground state d where $H_{\varepsilon,Z}^{\mathrm{TF}}$ is defined as in (54).

To this end we introduce the localizing functions

$$v_\kappa(\mathbf{r}) := \cos(\psi(|\mathbf{r} - \mathfrak{R}_\kappa|/R))$$

where $\psi(t)$ is some continuous, piecewise differentiable, monotone decreasing function which vanishes, if $t < 1/4$, and which is $\pi/2$, if $t > 1/2$. Note that the supports of these functions have at most finitely many points in common because R is the minimal nuclear distance. We also define

$$v_0 := \sqrt{1 - \sum_{\kappa=1}^K v_\kappa^2}.$$

Now pick the density $\varrho(\mathbf{r}) := \sum_{\kappa=1}^K \varrho_{Z_\kappa}^{\mathrm{TF}}(|\mathbf{r} - \mathfrak{R}_\kappa|)$ and denote the one-particle density matrix belonging to d by d_1 . Note that $\mathrm{tr} d_1 = N$. By the correlation inequality ([13], [11]) and the localization formula using the above decomposition of unity we have

$$\begin{aligned} (60) \quad & \mathrm{tr}(H_{N,\bar{Z},\bar{R}}^\varepsilon d) \\ & \geq \mathrm{tr} \left\{ \left[-\Delta - \sum_{\kappa=1}^K \varphi_{Z_\kappa}^{\mathrm{TF}}(\cdot - \mathfrak{R}_\kappa) - \varepsilon U_\lambda \right] d_1 \right\} \\ & \quad - D(\varrho, \varrho) + \sum_{\substack{\kappa, \kappa'=1 \\ \kappa < \kappa'}}^K \frac{Z_\kappa Z_{\kappa'}}{|\mathfrak{R}_\kappa - \mathfrak{R}_{\kappa'}|} - \mathrm{const} \lambda^{\frac{5}{3}} \\ & \geq \mathrm{tr} \left\{ \sum_{\kappa=0}^K v_\kappa \left[-\Delta - \sum_{\kappa'=1}^K \varphi_{Z_{\kappa'}}^{\mathrm{TF}}(|\cdot - \mathfrak{R}_{\kappa'}|) - \varepsilon U_\lambda \right] v_\kappa d_1 \right\} \\ & \quad - \left\| \sum_{\kappa=0}^K |\mathrm{grad} v_\kappa|^2 \right\|_\infty N - \sum_{\kappa=1}^K D(\varrho_{Z_\kappa}^{\mathrm{TF}}, \varrho_{Z_\kappa}^{\mathrm{TF}}) - \mathrm{const} \lambda^{\frac{5}{3}}. \end{aligned}$$

To obtain (60) we used the spherical symmetry of $\varphi_1, \dots, \varphi_K$ to show that

$$D(\varrho_{Z_\kappa}^{\mathrm{TF}}, \varrho_{Z_{\kappa'}}^{\mathrm{TF}}) \leq Z_\kappa Z_{\kappa'} |\mathfrak{R}_\kappa - \mathfrak{R}_{\kappa'}|^{-1}.$$

Now pick any arbitrary pair of different indices $\kappa, \kappa' \in \{1, \dots, K\}$. On the support of v_κ we have

$$\varphi_{Z_{\kappa'}}^{\mathrm{TF}}(|\mathbf{r} - \mathfrak{R}_{\kappa'}|) \leq \frac{2^2 3^4 \pi^2}{q^2 (R/2)^4}$$

where we use the fact that the Sommerfeld solution of the Thomas-Fermi equation is a pointwise upper bound of the Thomas-Fermi potential ([12], Section V.2). Thus on the support of v_κ we have

$$\sum_{\kappa'=1, \kappa'+\kappa}^K \phi_{Z_{\kappa'}}^{\text{TF}}(|\mathbf{r} - \mathfrak{R}_{\kappa'}|) \leq \frac{2^6 3^4 (K-1)}{q^2 R^4}.$$

For the derivative of the v_κ , which governs the localization error, we have the following uniform estimate

$$\sum_{\kappa=0}^K |\text{grad } v_\kappa|^2 = |\psi'(|\mathbf{r} - \mathfrak{R}_\kappa|)|^2 / R^2 = 4\pi^2 / R^2$$

where, for definiteness, we picked ψ to be the linear function interpolating between 0 and $\pi/2$ on the interval $[1/4, 1/2]$. Note that outside the annuli of thickness $R/2$ centered at the nuclei the derivatives vanish, in fact, whereas in these annuli the bound is actually an equality.

Next we show that there is no relevant contribution to the energy stemming from regions which are not close to at least one of the nuclei, i.e., from the support of v_0 . To this end we first remark that the supports of U_λ and v are disjoint for λ large enough. We set $Y := \text{supp } v_0 \cap (\text{supp } v_1 \cup \dots \cup v_K)$. Then, by the Lieb-Thirring inequality,

$$\begin{aligned} & \text{tr} \left\{ \left[v_0 \left(-\Delta - \sum_{\kappa=1}^K \phi_{Z_\kappa}^{\text{TF}} - \chi_Y \frac{4\pi^2}{R^2} \right) v_0 \right] d_1 \right\} \\ & \geq -\text{const} \left(\int_{|\mathbf{r}| > \frac{R}{4}} |\mathbf{r}|^{-10} d\mathbf{r} + \int_{\frac{R}{2} > |\mathbf{r}| > \frac{R}{4}} d\mathbf{r} \right) \geq -\text{const}(R^{-7} + R^3) \geq -\text{const } \lambda^{7/4}. \end{aligned}$$

This yields

$$\begin{aligned} & \text{tr}(H_{N, \tilde{Z}, \tilde{R}}^\varepsilon d) \\ & \geq \sum_{\kappa=1}^K \left\{ \text{tr} \left[-\Delta - \phi_{Z_\kappa}^{\text{TF}} - \frac{4\pi^2}{R^2} + \frac{2^6 3^4 (K-1)}{q^2 R^4} \chi_{B_{\frac{R}{2}}(0)} - \varepsilon \delta_{1,\kappa} \lambda^2 U(\lambda \cdot) d_1 \right] - D(q_{Z_\kappa}^{\text{TF}}, q_{Z_\kappa}^{\text{TF}}) \right\} - \text{const } \lambda^{7/4} \\ & \geq \sum_{\kappa=1}^K \left\{ \text{tr} \left(-\Delta - \phi_{Z_\kappa}^{\text{TF}} - \varepsilon \delta_{1,\kappa} U_\lambda \right) - D(q_{Z_\kappa}^{\text{TF}}, q_{Z_\kappa}^{\text{TF}}) \right\} - N \left(\frac{4\pi^2}{R^2} + \frac{2^6 3^4 (K-1)}{q^2 R^4} \right) - \text{const } \lambda^{7/4} \\ & \geq \inf \sigma(H_{\varepsilon, Z_1}^{\text{TF}}) - D(q_{Z_1}^{\text{TF}}, q_{Z_1}^{\text{TF}}) + \sum_{\kappa=2}^K [\inf \sigma(H_{0, Z_\kappa}^{\text{TF}}) - D(q_{Z_\kappa}^{\text{TF}}, q_{Z_\kappa}^{\text{TF}})] - \text{const} \left(\frac{\lambda}{R^4} + \lambda^{7/4} \right) \\ & \geq \inf \sigma(H_1^\varepsilon) - D(q_{Z_\kappa}^{\text{TF}}, q_{Z_\kappa}^{\text{TF}}) + \sum_{\kappa=2}^K [\inf \sigma(H_\kappa) - D(q_{Z_\kappa}^{\text{TF}}, q_{Z_\kappa}^{\text{TF}})] - \text{const } \lambda^{2-\delta} \end{aligned}$$

for some sufficiently small but positive δ . Combining this with Solovej's upper bound reduces the convergence question to that of the one-center case. $\square$

A. Appendix: Facts about the atomic ground state energy

According to [22] we have

$$(61) \quad E_{Z,Z} \leq E_{\text{TF}}(Z, Z) + \frac{q}{8} Z^2 + \text{const } Z^{\frac{47}{24}},$$

and according to [23] (see also [24] and Hughes [9])

$$(62) \quad \begin{aligned} E_{Z,Z} &\geq \sum_{l=0}^{L-1} q(2l+1) \text{tr}(H_{l,0,Z}^H)_- \\ &\quad + \sum_{l=L}^{\infty} q(2l+1) \text{tr}(H_{l,0,Z})_- - D(\varrho_{\text{TF}}, \varrho_{\text{TF}}) - \text{const } Z^{\frac{5}{3}} \\ &\geq E_{\text{TF}}(Z, Z) + \frac{q}{8} Z^2 - \text{const } Z^{\frac{17}{9}} \log Z \end{aligned}$$

with $L = [Z^{\frac{1}{9}}]$. Combining (61) and (62) gives

$$(63) \quad \begin{aligned} E_{Z,Z} &= \sum_{l=0}^{L-1} q(2l+1) \text{tr}(H_{l,0,Z}^H)_- \\ &\quad + \sum_{l=L}^{\infty} q(2l+1) \text{tr}(H_{l,0,Z})_- - D(\varrho_{\text{TF}}, \varrho_{\text{TF}}) + O(Z^{\frac{47}{24}}). \end{aligned}$$

References

- [1] *V. Bach*, A proof of Scott's conjecture for ions, *Rep. Math. Phys.* **28** (2) (1989), 213–248.
- [2] *C. Fefferman* and *L. Seco*, Eigenfunctions and eigenvalues of ordinary differential operators, *Adv. Math.* **95**(2) (1992), 145–305.
- [3] *C. Fefferman* and *L. Seco*, The density of a one-dimensional potential, *Adv. Math.* **107**(2) (1994), 187–364.
- [4] *C. Fefferman* and *L. Seco*, The eigenvalue sum of a one-dimensional potential, *Adv. Math.* **108**(2) (1994), 263–335.
- [5] *C. Fefferman* and *L. Seco*, On the Dirac and Schwinger corrections to the ground-state energy of an atom, *Adv. Math.* **107**(1) (1994), 1–188.
- [6] *C. Fefferman* and *L. Seco*, The density in a three-dimensional radial potential, *Adv. Math.* **111** (1) (1995), 88–161.
- [7] *C.L. Fefferman* and *L.A. Seco*, Aperiodicity of the Hamiltonian flow in the Thomas-Fermi potential, *Rev. Math. Iberoamer.* **9**(3) (1993), 409–551.
- [8] *O.J. Heilmann* and *E.H. Lieb*, Electron density near the nucleus of a large atom, *Phys. Rev. A* **52** (5) (1995), 3628–3643.
- [9] *W. Hughes*, An atomic lower bound that agrees with Scott's correction, *Adv. Math.* **79** (1990), 213–270.
- [10] *V.J. Ivrii* and *I.M. Sigal*, Asymptotics of the ground state energies of large Coulomb systems, *Ann. Math.* **138**(2) (1993), 243–335.
- [11] *E.H. Lieb*, Thomas-Fermi and related theories of atoms and molecules, *Rev. Mod. Phys.* **53**(4) (1981), 603–641.
- [12] *E.H. Lieb* and *B. Simon*, The Thomas-Fermi theory of atoms, molecules and solids, *Adv. Math.* **23** (1977), 22–116.
- [13] *E.H. Lieb* and *W.E. Thirring*, Bound for the kinetic energy of Fermions which proves the stability of matter, *Phys. Rev. Lett.* **35**(11) (1975), 687–689.

- [14] *F.W.J. Olver*, Error bounds for the Lioville-Green (or WKB) approximation, *Proc. Camb. Phil. Soc.* **57** (1961), 790–810.
- [15] *F.W.J. Olver*, *Tables for Bessel Functions of Moderate or Large Orders*, volume 6 of *Mathematical Tables*, Her Majesty's Stationary Office, London, 1 edition, 1962.
- [16] *F.W.J. Olver*, Bessel functions of integer order, in: Milton Abramowitz and Irene A. Stegun, editors, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, chapter 9, pages 355–433, Dover Publications, New York, 5 edition, 1968.
- [17] *F.W.J. Olver*, *Asymptotics and Special Functions*, Academic Press, New York, 1 edition, 1974.
- [18] *M. Reed* and *B. Simon*, *Methods of Modern Mathematical Physics*, volume 4: *Analysis of Operators*, Academic Press, New York, 1 edition, 1978.
- [19] *J.M.C. Scott*, The binding energy of the Thomas-Fermi atom, *Phil. Mag.* **43** (1952), 859–867.
- [20] *H. Siedentop*, An upper bound for the atomic ground state density at the nucleus, *Lett. Math. Phys.* **32**(3) (1994), 221–229.
- [21] *H. Siedentop*, Bound for the atomic ground state density at the nucleus, *CRM Proc. Lect. Notes* **8** (1995), 271–275.
- [22] *H. Siedentop* and *R. Weikard*, On the leading energy correction for the statistical model of the atom: Interacting case, *Comm. Math. Phys.* **112** (1987), 471–490.
- [23] *H. Siedentop* and *R. Weikard*, On the leading correction of the Thomas-Fermi model: Lower bound – with an appendix by A. M. K. Müller, *Invent. Math.* **97** (1989), 159–193.
- [24] *H. Siedentop* and *R. Weikard*, A new phase space localization technique with application to the sum of negative eigenvalues of Schrödinger operators, *Ann. Sc. Éc. Norm. Sup.* **24**(2) (1991), 215–225.
- [25] *Jan Philip Solovej*, In preparation.

Institut de Recherche Mathématique de Rennes, U.R.A. 305 – CNRS – Université de Rennes I,
F-35042 Rennes Cedex, France

Departments of Mathematics and Physics, Princeton University, Princeton, NJ 08544-0708, USA

Matematisk institutt, Universitetet i Oslo, Postboks 1053, N-0316 Oslo, Norway

Eingegangen 7. März 1995, in revidierter Fassung 21. August 1995

Asymptotics of Natural and Artificial Atoms in Strong Magnetic Fields*

Elliott H. Lieb
Princeton University

Jan Philip Solovej
Aarhus University

Jakob Yngvason
University of Iceland

1 Introduction

Magnetic fields in terrestrial experiments have only tiny effects on the ground-state properties of conventional atoms. The reason is that the natural atomic unit for magnetic field strength, $B_0 = m^2 e^3 c / \hbar^3 = 2.35 \times 10^5$ Tesla, is enormous compared with laboratory fields, which are seldom larger than 10 T. Here m denotes the electron mass, e the elementary charge, and $\hbar$ and c have their usual meaning. The unit B_0 is the field strength B at which the magnetic length $\ell_B = (\hbar c / (eB))^{1/2}$ ($\sim$ cyclotron radius for an electron in the lowest Landau level) is equal to the Bohr radius $a_0 = \hbar^2 / (me^2)$. Equivalently, at $B = B_0$ the Landau energy $\hbar\omega_B$, with $\omega_B = eB / (mc)$ the cyclotron frequency, becomes equal to the Rydberg energy e^2 / a_0 . For $B \ll B_0$ distortions of ground-state wave functions and energy level shifts due to the magnetic field will therefore be small, and their standard treatment by means of perturbation theory is completely adequate.

Magnetic fields comparable to and even much larger than B_0 , however, exist around cosmic bodies like white dwarfs and neutron stars [1]. In fields of such strength the magnetic forces are no longer a small perturbation of the Coulomb forces, and may drastically alter the structure of atoms and matter in bulk. The discovery of pulsars in 1967 spurred the interest of astrophysicists in the properties of atoms in high magnetic fields, and a considerable number of papers devoted to this subject have appeared since the early 1970's (see [2,3] for a history and a list of references.)

*This is the text of a lecture given by J. Yngvason at the XIth International Congress of Mathematical Physics, Paris 1994. It appeared originally in the proceedings of the congress, ed. by D. Iagolnitzer, pp. 185–205, International Press 1995, but has been updated and slightly expanded for the present publication. It summarizes the joint work [2], [3] and [5] of the three authors and is presented here in lieu of [2] and [3] which are too long to be included in this volume.

In recent years remarkably small two-dimensional structures, called *quantum dots*, have been fabricated by semiconductor technology. In quantum dots electrons, ranging in number from zero to several thousand, are confined within regions of diameter $\sim 10\text{--}1000$ nm. Quantum dots have many properties in common with natural atoms and can justly be regarded as two-dimensional *artificial atoms* (see [4] for a review). Because the electrons interact with the semiconductor crystal their effective parameters can differ markedly from the free values. Thus the effective mass m_* in GaAs is approximately $0.07m$ and the effective charge $e_* = e/\sqrt{\epsilon} \approx 0.3e$, where $\epsilon \approx 13$ is the dielectric constant. Corresponding to these values there is an effective Bohr radius, $a_* = \hbar^2/(m_*e_*^2) \approx 185a_0$. The magnetic length ℓ_B becomes equal to a_* for $B = B_* = (a_0/a_*)^2 B_0 \approx 7$ T, and hence effects that for natural atoms require magnetic fields of astronomical strength, can for artificial atoms be studied in any well-equipped laboratory.¹

In the papers [2,3,5] (see also [6–8]) we have studied the quantum mechanical ground states of natural and artificial atoms in homogeneous magnetic fields of arbitrary strength. Our main results are limit theorems for the ground-state energy and electronic density as the number of electrons, N , and the strength of the attractive potential, measured by the nuclear charge Z (natural atoms) or a coupling constant K (quantum dots), tend to infinity with N/Z or N/K fixed. The ground states can in this limit be evaluated *exactly* by five nonlinear functionals for natural atoms and three for quantum dots, corresponding to different physics at different scales of the magnetic field B as measured by powers of N . The asymptotic theories are amenable to computer studies and results of numerical computations carried out by K. Johnsen and Ö. Rönkvallsson will be presented below.

Owing to their higher dimensionality, natural atoms have a richer structure than quantum dots, and the larger part of this review is devoted to the former. Parts of the analysis are similar for both cases and after having discussed natural atoms one can be more brief about the dots. There are some marked differences between the two cases, however. In particular, the repulsive interaction of the electrons in a quantum dot is the three-dimensional Coulomb potential, but the motion of the electrons is two-dimensional. This gives rise to somewhat peculiar electrostatics, since Newton's law for the field produced by a spherically symmetric charge distribution no longer applies. Also, the estimate of the kinetic energy in terms of a functional of the density is harder in 2D than in 3D, and for strong fields the treatment of the indirect part of the Coulomb interaction requires different techniques in the two cases.

¹Natural atoms in highly *excited* states (“Rydberg atoms”) can also be strongly affected by magnetic fields of just a few T, giving rise to various forms of “quantum chaos”. This subject will not be discussed here since we deal exclusively with ground-state properties.

2 Heavy Atoms in High Magnetic Fields

2.1 The Hamiltonian

The starting point for the investigation of natural, three-dimensional atoms is the Hamiltonian for N electrons in the Coulomb field of a nucleus with charge Ze and a homogeneous magnetic field $\mathbf{B} = (0, 0, B)$:

$$H_{N,B,Z} = \sum_{i=1}^N \left\{ [(\mathbf{p}^{(i)} + \mathbf{A}(x^{(i)})) \cdot \sigma^{(i)}]^2 - Z|x^{(i)}|^{-1} \right\} + \sum_{1 \leq i < j \leq N} |x^{(i)} - x^{(j)}|^{-1} \quad (1)$$

Here $\mathbf{A}(x) = (1/2)(-x_2 B, x_1 B, 0)$ with $x = (x_1, x_2, x_3) \in \mathbf{R}^3$ is the vector potential, $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ is the vector of Pauli matrices, and the units are chosen such that $\hbar = e = 2m_e = 1$, $c = 1/\alpha \approx 137$. The Hamiltonian $H_{N,B,Z}$ operates on antisymmetric wave functions $\Psi \in \bigwedge_1^N L^2(\mathbf{R}^3; \mathbf{C}^2)$ of space and spin variables. The ground-state energy is

$$E^Q(N, B, Z) = \inf_{(\Psi, \Psi)=1} (\Psi, H_{N,B,Z} \Psi), \quad (2)$$

and the ground-state electron density

$$\rho_{N,B,Z}^Q(x) = N \sum_{s^{(i)}=\pm \frac{1}{2}} \int \left| \Psi_0(x, x^{(2)}, \dots, x^{(N)}; s^{(1)}, \dots, s^{(N)}) \right|^2 dx^{(2)} \dots dx^{(N)}, \quad (3)$$

where Ψ_0 is a ground-state wave function. We study the asymptotic behavior of the energy and density as $N, Z \rightarrow \infty$ with N/Z fixed. The magnetic field B is allowed to vary with N as well. If $B = 0$ the leading contribution is well known ([9,10]): It is given by Thomas–Fermi (TF) theory, which for $Z > 20$ is accurate within a few percent as far as the bulk of the electrons is concerned. The theories described here are the generalizations to the case that $B \rightarrow \infty$ as $N, Z \rightarrow \infty$. This case is important for the study of neutron star surfaces, which are believed to consist mostly of iron atoms ($Z = 26$) in magnetic fields as large as $10^7 - 10^9$ T.

Before discussing the asymptotics of (2) and (3) let us recall the basic facts about the one-particle “kinetic energy” operator (including the interaction of the electron magnetic moment with the field)

$$H_A = [(\mathbf{p} + \mathbf{A}(x)) \cdot \sigma]^2 = (\mathbf{p}_\perp + \mathbf{A}(x))^2 + B\sigma_3 + p_3^2. \quad (4)$$

Here $\mathbf{p}_\perp$ and p_3 denote respectively the momentum perpendicular and parallel to the magnetic field. The spectrum of H_A decomposes into Landau bands $\nu = 0, 1, \dots$ with energy range $[2\nu B, \infty[$, and we have

$$\Pi_\nu H_A \Pi_\nu = -\partial_3^2 + 2\nu B \quad (5)$$

where Π_ν is the projector on the eigenfunctions in the ν th band, and $\partial_3 := \partial/\partial x_3$. The degeneracy of the Landau level with energy $2\nu B$ is, per unit area perpendicular to the field, $d_0(B) = (\pi\ell_B^2)^{-1} = B/(2\pi)$ for $\nu = 0$ and $d_\nu(B) = 2(\pi\ell_B^2)^{-1} = B/\pi$ for $\nu \geq 1$.

2.2 The Five Asymptotic Regions

In the following the symbols $\ll$, $\sim$ and $\gg$ are meant to indicate the asymptotic behavior of N -dependent quantities as $N \rightarrow \infty$. Thus $B \ll Z^p$, $B \sim Z^p$, and $B \gg Z^p$ mean, respectively, that $B/Z^p \rightarrow 0$, $B/Z^p \rightarrow (\text{const.}) \neq 0$ and $B/Z^p \rightarrow \infty$ as $N \rightarrow \infty$ with $Z \sim N$. As already noted in [11–14] the exponents $p = 4/3$ and $p = 3$ are of special importance for the asymptotic properties of the ground states of (1). The analysis in [2] and [3] distinguishes the following parameter regions and establishes their precise status:

$$\begin{array}{lll} \text{Region 1: } B \ll Z^{4/3} & \text{Region 2: } B \sim Z^{4/3} & \text{Region 3: } Z^{4/3} \ll B \ll Z^3 \\ \text{Region 4: } B \sim Z^3 & \text{Region 5: } B \gg Z^3 & \end{array}$$

A first orientation about the different physics in the different regions may be obtained by the following heuristic reasoning.

By the Pauli principle each electron can be thought of as occupying a “private room”. At $B = 0$ its spatial extension a is independent of direction and the kinetic energy is $\varepsilon_{\text{kin}} \sim 1/a^2$. To minimize the Coulomb energy due to the attraction of the nucleus the electrons arrange themselves in a sphere around the nucleus. If R denotes the approximate radius of the sphere the potential energy of an electron is $\varepsilon_{\text{pot}} \sim -Z/R \sim -Z/(N^{1/3}a)$ since $Na^3 \sim R^3$. Optimizing $\varepsilon = \varepsilon_{\text{kin}} + \varepsilon_{\text{pot}}$ with respect to a gives

$$a \sim Z^{-1}N^{1/3} \sim Z^{-2/3} \quad \text{and} \quad R \sim Z^{-1}N^{2/3} \sim Z^{-1/3}. \quad (6)$$

The energy per electron, ε , and the ground-state energy, $E = N\varepsilon$, are

$$\varepsilon \sim -Z^2N^{-2/3} \sim -Z^{4/3} \quad \text{and} \quad E \sim -Z^2N^{1/3} \sim -Z^{7/3}. \quad (7)$$

The repulsive energy, $\varepsilon_{\text{rep}} \sim N/R$, has been ignored since it does not affect the rough estimates of energies and sizes presented here, although it is of course crucial for other questions, e.g., the maximal number of electrons that a nucleus can bind.

Since the effects of a magnetic field on the energy are of the order B , one expects only a small perturbation of the $B = 0$ picture as long as $B \ll |\varepsilon|$, i.e., $B \ll Z^{4/3}$. The wave functions are little affected by the magnetic field, provided $a \ll \ell_B$. Since $a \sim Z^{-2/3}$ and $\ell_B \sim B^{-1/2}$, this leads to the same condition as before. On the other hand, for $B \sim Z^{4/3}$ the distance between electrons is comparable with the magnetic length and the field will start to have effects to leading order. These contributions will involve all Landau levels.

At $B \gg Z^{4/3}$ energy differences between Landau levels are much larger than the Coulomb energy scale, and the electrons will essentially be confined to the lowest Landau band. The private room of an electron is now a cylinder with

radius $\sim \ell_B \sim B^{-1/2}$ and length $L \gg \ell_B$. The kinetic energy is $\varepsilon_{\text{kin}} \sim 1/L^2$. A spherically symmetric arrangement of the electrons around the nucleus is still optimal and possible, provided the condition $L \ll R$, with R the radius of the sphere, is compatible with $N\ell_B^2 \sim R^3$ and $1/L^2 - Z/R = \text{minimum}$. The last two conditions give

$$R \sim Z^{-1/5} N^{2/5} B^{-2/5} \sim Z^{1/5} B^{-2/5} = Z^{-1/3} (B/Z^{4/3})^{-2/5} \quad (8)$$

and hence

$$L \sim Z^{-3/5} N^{1/5} B^{-1/5} \sim ((BZ^2/N^5)^{1/5}) R \sim (B/Z^3)^{1/5} R. \quad (9)$$

The condition for sphericity, $L \ll R$, thus becomes $B \ll Z^3$. The energy is

$$E \sim -Z^{6/5} N^{3/5} B^{2/5} \sim -Z^{9/5} B^{2/5} = -Z^{7/3} (B/Z^{4/3})^{2/5}. \quad (10)$$

At $B \sim Z^3$ the atom begins to deviate from spherical shape because the condition $L \ll R$ no longer holds for the optimal L .

If $B \gg Z^3$ the extension of an atom along the field is the same as for one electron, i.e., the atom is approximately a cylinder of length L and radius

$$R \sim (N\ell_B^2)^{1/2} \sim (N/B)^{1/2} \ll L. \quad (11)$$

The Coulomb energy of an electron is now $\varepsilon_{\text{pot}} \sim -(Z/L) \ln(L/R)$ and the optimum of $\varepsilon = \varepsilon_{\text{kin}} + \varepsilon_{\text{pot}}$ with $\varepsilon_{\text{kin}} \sim 1/L^2$ is obtained for

$$L \sim Z^{-1} [\ln(B/(Z^2 N))]^{-1} \sim Z^{-1} [\ln(B/(Z^3))]^{-1}. \quad (12)$$

We see that $L/R \sim (B/(Z^2 N))^{1/2} \ln(B/(Z^2 N))$, so we indeed have $L \gg R$ for $B \gg Z^3$. The energy $E = N\varepsilon$ is

$$E \sim -Z^2 N \ln[(B/(Z^2 N))] \sim -Z^3 \ln(B/Z^3). \quad (13)$$

To summarize, the heuristic considerations have disclosed the asymptotic regions $B \ll Z^{4/3}$, $Z^{4/3} \ll B \ll Z^3$ and $B \gg Z^3$, where the simple expressions (7), (10) and (13) for the ground-state energy are plausible, and the transition regions $B \sim Z^{4/3}$ and $B \sim Z^3$, where a more complex behavior is expected.

It is a long way from the rough estimates given above to rigorous theorems about the ground-state asymptotics of (1). Clearly the cases $B \sim Z^{4/3}$ and $B \sim Z^3$ are the most challenging, and in fact the other three cases can be regarded as limits of these two. If $B \sim Z^{4/3}$ the asymptotics turns out to be given by a B -dependent density functional theory of the TF type, that was first introduced by Tomishima and Yonei [15]. Its limit for $B/Z^{4/3} \rightarrow 0$ is the usual $B = 0$ TF theory, whereas for $B/Z^{4/3} \rightarrow \infty$, but $B/Z^3 \rightarrow 0$, it passes into another TF-type theory first considered by Kadomtsev [11]. For $B \sim Z^3$ a new type of functional, depending on density *matrices* rather than the density alone gives the correct limit. If $B/Z^3 \rightarrow \infty$ it simplifies to a one-dimensional theory that is solvable *in closed form*.

2.3 The Semiclassical Theories

The density functional that correctly describes the ground state for $B \sim Z^{4/3}$ is the *magnetic Thomas–Fermi functional* (MTF)

$$\mathcal{E}^{\text{MTF}}[\rho; B, Z] = \int \tau_B(\rho(x))dx - Z \int |x|^{-1}\rho(x)dx + D(\rho, \rho). \quad (14)$$

Here $\rho \geq 0$ is the electron density,

$$D(\rho, \rho) = \frac{1}{2} \iint \rho(x)\rho(y)|x - y|^{-1}dxdy \quad (15)$$

is the Coulomb repulsion and τ_B is the kinetic energy density of a free-electron gas in magnetic field B . This function is the Legendre transform of the pressure P_B and is most conveniently characterized by the fact that $\tau_B(0) = 0$ and the derivative τ'_B is the inverse of the derivative of the pressure, i.e.,

$$P'_B(\tau'_B(\rho)) = \rho \quad (16)$$

with

$$P'_B(w) = (2\pi^2)^{-1}B[w^{1/2} + 2 \sum_{\nu \geq 1} [w - 2B\nu]_+^{1/2}]. \quad (17)$$

The function $\tau_B(\rho)$ is for large ρ bounded above by $(\text{const.})\rho^{5/3}$, and for $B \rightarrow 0$ it tends to the kinetic energy density at $B = 0$, $\tau_0(\rho) = (3/5)(3\pi^2)^{2/3}\rho^{5/3}$. Standard TF theory, where τ_B is replaced by τ_0 , is thus a limiting case of (14). In sufficiently strong fields only the contribution from the lowest Landau level in (17) is relevant, and τ_B is replaced by

$$\tau_B^{\text{STF}}(\rho) = (4\pi^4/3)\rho^3/B^2. \quad (18)$$

The corresponding theory will be referred to as the STF theory; it was first studied in [11,14,16].

For each N there is a minimal energy $E^{\text{MTF}}(N, B, Z)$ and a unique minimizing density $\rho_{N,B,Z}^{\text{MTF}}$ for the functional (14):

$$E^{\text{MTF}}(N, B, Z) = \inf \{ \mathcal{E}^{\text{MTF}}[\rho] \mid \int \rho = N \} = \mathcal{E}^{\text{MTF}}[\rho_{N,B,Z}^{\text{MTF}}]. \quad (19)$$

This is proved by the methods of [9] and [10]. Corresponding quantities for the TF and STF theories are denoted $E^{\text{TF}}(N, Z)$, $\rho_{N,Z}^{\text{TF}}$, $E^{\text{STF}}(N, B, Z)$ and $\rho_{N,B,Z}^{\text{STF}}$. The following scaling laws for the energies hold with $\lambda = N/Z$, $\beta = B/Z^{4/3}$:

$$E^{\text{TF}}(N, Z) = Z^{7/3}E^{\text{TF}}(\lambda, 1) \quad (20)$$

$$E^{\text{MTF}}(N, B, Z) = Z^{7/3}E^{\text{MTF}}(\lambda, \beta, 1) \quad (21)$$

$$E^{\text{STF}}(N, B, Z) = Z^{7/3}\beta^{2/5}E^{\text{STF}}(\lambda, 1, 1) \quad (22)$$

The corresponding relations for the densities are

$$\rho_{N,Z}^{\text{TF}}(x) = Z^2 \rho_{\lambda,1}^{\text{TF}}(Z^{1/3}x) \quad (23)$$

$$\rho_{N,B,Z}^{\text{MTF}}(x) = Z^2 \rho_{\lambda,\beta,1}^{\text{MTF}}(Z^{1/3}x) \quad (24)$$

$$\rho_{N,B,Z}^{\text{STF}}(x) = Z^2 \beta^{6/5} \rho_{\lambda,1,1}^{\text{STF}}(Z^{1/3} \beta^{2/5}x). \quad (25)$$

Thus the TF theory has one nontrivial parameter, λ , MTF has two parameters, λ and β , and STF again only one, λ .

The basic quantum mechanical limit theorems for the three regions $B \ll Z^{4/3}$, $B \sim Z^{4/3}$ and $Z^{4/3} \ll B \ll Z^3$ are

Theorem 2.1 (Energy asymptotics in Regions 1–3) *Let $N, Z \rightarrow \infty$ with N/Z fixed. If $B/Z^3 \rightarrow 0$, then*

$$E^{\text{Q}}(N, B, Z)/E^{\text{MTF}}(N, B, Z) \rightarrow 1. \quad (26)$$

If $B/Z^{4/3} \rightarrow 0$, then

$$E^{\text{Q}}(N, B, Z)/E^{\text{TF}}(N, Z) \rightarrow 1. \quad (27)$$

Finally, if $B/Z^{4/3} \rightarrow \infty$, but $B/Z^3 \rightarrow 0$, then

$$E^{\text{Q}}(N, B, Z)/E^{\text{STF}}(N, B, Z) \rightarrow 1. \quad (28)$$

Theorem 2.2 (Density asymptotics in Regions 1–3) *Let $N, Z \rightarrow \infty$ with $\lambda = N/Z$ and $\beta = B/Z^{4/3}$ fixed. Then*

$$Z^{-2} \rho^{\text{Q}}(Z^{-1/3}x) \rightarrow \rho_{\lambda,\beta,1}^{\text{MTF}}(x). \quad (29)$$

If $B/Z^{4/3} \rightarrow 0$,

$$Z^{-2} \rho^{\text{Q}}(Z^{-1/3}x) \rightarrow \rho_{\lambda,1}^{\text{TF}}(x). \quad (30)$$

If $B/Z^{4/3} \rightarrow \infty$, but $B/Z^3 \rightarrow 0$, then

$$Z^{-2} \beta^{-6/5} \rho^{\text{Q}}(Z^{-1/3} \beta^{-2/5}x) \rightarrow \rho_{\lambda,1,1}^{\text{STF}}(x). \quad (31)$$

The limits are in the sense of weak convergence in $L_{\text{loc}}^{5/3}(\mathbf{R}^3)$.

Theorems 2.1–2 are proved in [3]. A brief discussion of the ingredients for the proofs will be given in Sect. 2.6. See also [17] for a special case and [18], [19] for related results and partial generalizations.

The theorems allow us to draw various conclusions about basic properties of heavy atoms in Regions 1–3, i.e., for $B \ll Z^3$. First, the theorems confirm the heuristic arguments given in Sect. 2.2 that in these parameter regions *atoms are spherical* to leading order with radius $\sim Z^{-1/3}(1 + (B/Z^{4/3}))^{-2/5}$. Second, it is a general property of TF-type theories that the maximum number of electrons that can be bound is $N = Z$. In other words, *negative ions do not exist* in

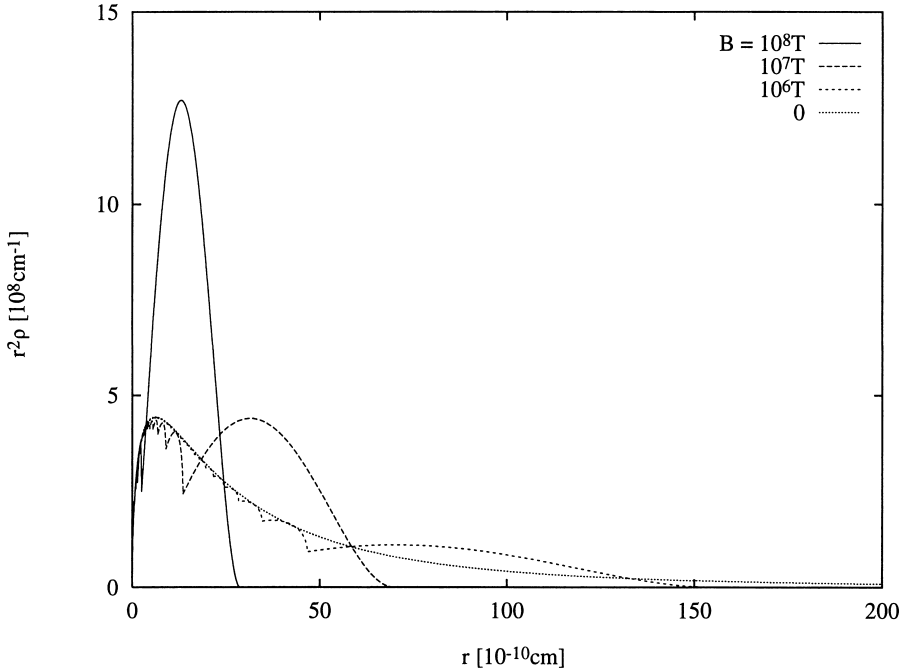


Figure 1: Plots of $|x|^2 \rho(x)$ in MTF theory for iron atoms in magnetic fields $B = 0, 10^6, 10^7, 10^8$ T.

the limit considered. Third, *molecules do not bind* in TF-type theories [9,10] which in the present context means that in Regions 1–3 the quantum mechanical binding energies of molecules are of lower order than ground state energies of atoms, which are of the order $Z^{7/3}(1 + B/Z^{4/3})^{2/5}$.

For further discussions of MTF theory and its astrophysical applications see [20,21] and references quoted therein. Figure 1, computed by Ö. Rönvaldsson, shows the density ρ^{MTF} for various values of B .

2.4 The Density Matrix Theory

In Region 4, i.e., for $B \sim Z^3$, atoms cease to be spherical and a new type of functional takes over from the semiclassical theories of the last subsection. The variable of this functional is not a density but a mapping $x_\perp \rightarrow \Gamma_{x_\perp}$ from $x_\perp = (x_1, x_2) \in \mathbf{R}^2$ into density matrices, i.e., nonnegative trace class operators on $L^2(\mathbf{R}, dx_3)$, satisfying the condition

$$0 \leq \Gamma_{x_\perp} \leq (B/(2\pi))I \quad (32)$$

for all $x_\perp$. Let $\Gamma_{x_\perp}(x_3, y_3)$ denote the integral kernel of $\Gamma_{x_\perp}$ and put $\rho_\Gamma(x) = \Gamma_{x_\perp}(x_3, x_3)$. The *Density matrix functional* (DM) is defined as follows:

$$\begin{aligned} \mathcal{E}^{\text{DM}}[\Gamma; B, Z] = & \int \left[-\partial_3^2 \Gamma_{x_\perp}(x_3, y_3) \right]_{x_3=y_3} dx_3 dx_\perp \\ & - Z \int |x|^{-1} \rho_\Gamma(x) dx + D(\rho_\Gamma, \rho_\Gamma). \end{aligned} \quad (33)$$

The corresponding energy $E^{\text{DM}}(N, B, Z)$ and minimizer $\Gamma_{N,B,Z}^{\text{DM}} \equiv \Gamma^{\text{DM}}$ (which can be shown to exist and be unique) are defined by

$$E^{\text{DM}}(N, B, Z) = \inf \{ \mathcal{E}^{\text{DM}}[\Gamma] : \int \rho_\Gamma \leq N \} = \mathcal{E}^{\text{DM}}[\Gamma^{\text{DM}}]. \quad (34)$$

To understand the motivations for (33) and the condition (32) recall first that for $B \gg Z^{4/3}$ the electrons in an approximate ground state Ψ of (1) are expected to be confined to the lowest Landau band. One may then, with small error, replace the kinetic energy operator (4) by $\Pi_0 H_A \Pi_0 = -\partial_3^2$; cf. (5). Also, one expects that for large N and Z the exchange-correlation part of the Coulomb repulsion is small compared with the direct part $D(\rho_\Psi, \rho_\Psi)$, where $\rho_\Psi(x)$ denotes the density corresponding to Ψ . Defining

$$\begin{aligned} \Gamma_{x_\perp}^\Psi(x_3, y_3) = & N \int \overline{\Psi(x_\perp, x_3; x^{(2)}, \dots, x^{(N)})} \\ & \times \Psi(x_\perp, y_3; x^{(2)}, \dots, x^{(N)}) dx^{(2)} \dots dx^{(N)} \end{aligned} \quad (35)$$

(summation over spins is understood) we have $\rho_\Psi(x) = \Gamma_{x_\perp}^\Psi(x_3, x_3)$. Thus,

$$(\Psi, H_{N,B,Z} \Psi) \approx \mathcal{E}^{\text{DM}}[\Gamma^\Psi; B, Z]. \quad (36)$$

The condition (32) can be traced to the fact that the density of states per unit area in the lowest Landau level is $d_0(B) = B/(2\pi)$. If Ψ is composed of wave functions in the lowest Landau band it is easy to deduce from this that Γ^Ψ must satisfy (32).

The DM functional (33) treats the electrostatic interactions classically but the kinetic energy for the motion along the magnetic field quantum mechanically by the $-\partial_3^2$ term. In directions perpendicular to the field the motion is restricted by the “hard core” condition (32). Note that without (32) the functional would not be bounded below.

The DM theory has two parameters, $\lambda = N/Z$ and $\eta = B/Z^3$. This is manifested in the scaling relations

$$E^{\text{DM}}(N, B, Z) = Z^3 E^{\text{DM}}(\lambda, \eta, 1) \quad (37)$$

and

$$\rho_{N,B,Z}^{\text{DM}}(x) = Z^4 \rho_{\lambda,\eta,1}^{\text{DM}}(Zx), \quad (38)$$

where ρ^{DM} is the density corresponding to the minimizer Γ^{DM} . The $\eta \rightarrow \infty$ limit will be considered in the next section.

The main quantum mechanical limit theorems for strong fields are

Theorem 2.3 (Energy asymptotics in Regions 3–5) *Let $N, Z \rightarrow \infty$ with N/Z fixed. If $B/Z^{4/3} \rightarrow \infty$, then*

$$E^Q(N, B, Z)/E^{\text{DM}}(N, B, Z) \rightarrow 1. \quad (39)$$

Theorem 2.4 (Density asymptotics in Region 4) *Let N, Z and $B \rightarrow \infty$ with $N/Z = \lambda$ and $B/Z^3 = \eta$ fixed. Then*

$$Z^{-4} \rho_{N,B,Z}^Q(Z^{-1}x) \rightarrow \rho_{\lambda,\eta,1}^{\text{DM}}(x) \quad (40)$$

weakly in L^1_{loc} .

Note that Theorem 2.3 overlaps with Theorem 2.1 since Region 3 is covered by both. A limit theorem for the density in Region 5 will be formulated in the next subsection.

Because of Theorems 2.3–4, DM theory is relevant for heavy atoms such as iron ($Z = 26$) in fields of neutron star strength. (At $B = 10^9$ T one has $\beta \approx 14$, $\eta \approx 0.06$ if $Z = 26$.) Numerical studies of the DM theory have been carried out by K. Johnsen [22], and Fig. 2 shows contour plots of the density ρ^{DM} for $N = Z = 26$ and three values of B : 10^8 , 10^9 and 10^{10} T. As apparent from these figures, the atomic core is still approximately spherical at the weakest field but the atom shrinks and becomes increasingly elongated as the field goes up. This is in accord with the order of magnitude estimates (11) and (12). Another noteworthy point is that for the strongest field the shape of the atom is simpler than for the weaker field, where the atom is composed of several cylindrical shells. The reason for this is that minimizers for the DM functional are obtained by seeking at each $x_\perp$ the lowest eigenfunctions for a one-dimensional Schrödinger Hamiltonian $-\partial_3^2 + V_{x_\perp}^{\text{DM}}(x_3)$ where $V_{x_\perp}^{\text{DM}}$ is the self-consistent potential generated by the nucleus and ρ^{DM} . The number of the cylindrical shells reflects the number of eigenfunctions that contribute to Γ^{DM} . At $B = 10^{10}$ T this number has dropped to one, in which case $\Gamma_{x_\perp}^{\text{DM}}(x_3, y_3) = \sqrt{\rho^{\text{DM}}(x)} \sqrt{\rho^{\text{DM}}(y)}$ and the DM theory simplifies to a density functional theory. The corresponding functional, denoted by $\mathcal{E}^{\text{SS}}$ (SS stands for *super-strong*), is

$$\mathcal{E}^{\text{SS}}[\rho; B, Z] = \int [\partial_3 \sqrt{\rho}]^2 dx - Z \int |x|^{-1} \rho(x) dx + D(\rho, \rho) \quad (41)$$

and condition (32) becomes

$$\int \rho(x_\perp, x_3) dx_3 \leq B/(2\pi) \quad (42)$$

for all $x_\perp \in \mathbf{R}^2$. We denote the ground-state energy of (41) by $E^{\text{SS}}(N, B, Z)$. The transition from DM to SS occurs at a certain critical value $\eta_c(\lambda)$ of the parameter $\eta = B/Z^3$, i.e., $E^{\text{DM}}(N, B, Z) = E^{\text{SS}}(N, B, Z)$ for $\eta \geq \eta_c$. Numerical studies [22] give the value $\eta_c = 0.148$ at $\lambda = 1$. This corresponds to $B = 2.44 \times 10^9$ T for iron atoms.

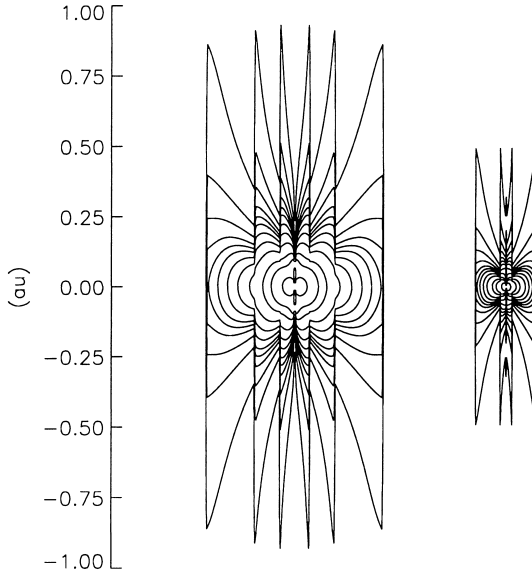


Figure 2: Contour plots of iron atoms in DM theory in magnetic fields $B = 10^8$, 10^9 and 10^{10} T.

In contrast to the semiclassical theories of Sect. 2.3 *negative ions and molecular binding are possible* in DM theory. In fact, in the extreme limit $\eta \rightarrow \infty$ considered in the next section the maximum value of λ is $\lambda_c = 2$, and the binding energy of a diatomic molecule is six times as large as the ground-state energy of individual atoms! At $B = 10^8$ T and $Z = 26$ the computations in [22] give $\lambda_c \approx 1.04$ and a ratio of about 0.18 between the binding energy of a diatomic molecule and the ground-state energy of an atom.

2.5 The Limit of Hyper-Strong Fields

According to (11) and (12) the length of the atom is expected to scale as $[Z \ln(B/Z^3)]^{-1}$ and the diameter as $(Z/B)^{-1/2}$ when $\eta = B/Z^3 \rightarrow \infty$. This suggests to change variables in the functional $\mathcal{E}^{\text{SS}}$ in accordance with these length scales and look for the $\eta \rightarrow \infty$ limit of the resulting functional. Closer examination of $\mathcal{E}^{\text{SS}}$ [22] shows that convergence is more rapid if $\ln \eta$ is replaced by another function with the same asymptotic behavior, namely the solution

$L(\eta)$ of the equation

$$(\eta/2)^{1/2} = L(\eta) \sinh(L(\eta)/2). \quad (43)$$

The rescaled Coulomb potential becomes

$$V_\eta(x) = L(\eta)^{-1} (\eta^{-1} L(\eta)^2 x_\perp^2 + x_3^2)^{-1/2} \quad (44)$$

and the rescaled SS functional is defined as

$$\mathcal{E}_\eta^{\text{SS}}(\rho) = \int [\partial_3 \sqrt{\rho}]^2 dx - \int V_\eta(x) \rho(x) dx + \int V_\eta(x-y) \rho(x) \rho(y) dx dy. \quad (45)$$

The energy

$$E_\eta^{\text{SS}}(\lambda) = \inf \{ \mathcal{E}_\eta^{\text{SS}}(\rho) \mid \int \rho dx \leq \lambda, \int \rho dx_3 \leq 1 \} \quad (46)$$

is related to E^{SS} by

$$E^{\text{SS}}(N, B, Z) = Z^3 L(\eta)^2 E_\eta^{\text{SS}}(\lambda). \quad (47)$$

Now $V_\eta(x) \rightarrow \delta(x_3)$ in distributional sense as $\eta \rightarrow \infty$. Defining $\bar{\rho}(x_3) = \int \rho(x_\perp, x_3) dx_\perp$ and the *Hyper-strong density functional* by

$$\mathcal{E}^{\text{HS}}[\bar{\rho}] = \int [\partial_3 \sqrt{\bar{\rho}}]^2 dx_3 - \bar{\rho}(0) + \int \bar{\rho}(x_3)^2 dx_3 \quad (48)$$

it follows that $\mathcal{E}_\eta^{\text{SS}}[\rho] \rightarrow \mathcal{E}^{\text{HS}}[\bar{\rho}]$ for all ρ . With a little more effort one shows that $E_\eta^{\text{SS}}(\lambda) = E^{\text{HS}}(\lambda) + O(L(\eta)^{-1})$, with $E^{\text{HS}}(\lambda)$ the ground-state energy of $\mathcal{E}^{\text{HS}}$, and thus by (47)

$$E^{\text{SS}}(N, B, Z) = Z^3 L(\eta)^2 E^{\text{HS}}(\lambda) + Z^3 O(L(\eta)). \quad (49)$$

Since $L(\eta)$ increases only logarithmically with η the convergence of SS to HS is fairly slow.

The hyper-strong functional is a rough approximation of the SS functional at all but extremely large η , but it has the interesting property that it can be minimized in closed form. One finds

$$E^{\text{HS}}(\lambda) = -\frac{1}{4}\lambda + \frac{1}{8}\lambda^2 - \frac{1}{48}\lambda^3 \quad (50)$$

for $\lambda \leq 2$ and $E^{\text{HS}}(\lambda) = E^{\text{HS}}(2) = -1/6$ for $\lambda \geq 2$. Writing the minimizer $\bar{\rho}^{\text{HS}}(x_3)$ as $\psi^{\text{HS}}(x_3)^2$ one has

$$\psi^{\text{HS}}(x) = \frac{\sqrt{2}(2-\lambda)}{4 \sinh[\frac{1}{4}(2-\lambda)|x| + c]} \quad \text{for } \lambda < 2 \quad (51)$$

$$\psi^{\text{HS}}(x) = \sqrt{2}(2+|x|)^{-1} \quad \text{for } \lambda \geq 2 \quad (52)$$

where $\tanh c = (2-\lambda)/2$.

Combining Theorem 2.3 with (49) one obtains

Theorem 2.5 (Energy asymptotics in Region 5) *Let $N, Z \rightarrow \infty$ with $N/Z = \lambda$ fixed. If $\eta = B/Z^3 \rightarrow \infty$, then*

$$E^Q(N, B, Z)/(Z^3 L(\eta)^2 E^{\text{HS}}(\lambda)) \rightarrow 1 \quad (53)$$

where $L(\eta)$ and $E^{\text{HS}}(\lambda)$ are given by (43) and (50).

There is also a limit theorem for the density:

Theorem 2.6 (Density asymptotics in Region 5) *Let $N, Z \rightarrow \infty$ with $N/Z = \lambda$ fixed. If $\eta = B/Z^3 \rightarrow \infty$ then*

$$[Z^2 L(\eta)]^{-1} \int \rho(x_\perp, [ZL(\eta)]^{-1} x_3) dx_\perp \rightarrow \bar{\rho}^{\text{HS}}(x_3) \quad (54)$$

weakly in $L^1_{\text{loc}}(\mathbf{R}, dx_3)$.

From these limit theorems and the explicit solution (50)-(52) of the HS theory one also obtains results about enhanced negative ionization and molecular binding of atoms in hyperstrong magnetic fields.

Theorem 2.7 (Negative ionization) *The maximal number, N_c^Q , of electrons that can be bound to an atom of nuclear charge Z , defined by*

$$N_c^Q = \max\{N : E^Q(N, B, Z) < E^Q(N - 1, B, Z)\}, \quad (55)$$

satisfies

$$\liminf N_c^Q/Z \geq 2 \quad (56)$$

as $Z \rightarrow \infty$ and $B/Z^3 \rightarrow \infty$.

The result about molecular binding is based on the fact that the nuclei of molecules coalesce in the HS limit, despite the nuclear repulsion. The striking fact is that *the binding energy is of the order of the atomic energy itself*—contrary to the ‘ordinary’ situation in which it is only a tiny perturbation. To state this precisely, let $E^Q(N, B, Z_1, \dots, Z_K)$ denote the quantum mechanical ground-state energy of a molecule of K nuclei with charges $Z_1, \dots, Z_K$ and N electrons. This energy includes the nuclear repulsion energy $\sum_{k < \ell} Z_k Z_\ell |R_k - R_\ell|^{-1}$, where the R_k denote the positions of the nuclei (that are assumed to be infinitely heavy), and it is understood that the infimum over all nuclear positions has been taken. We have

Theorem 2.8 (Bound atoms are isocentric for large B) *Put $Z_{\text{tot}} = Z_1 + \dots + Z_K$. If $B/Z_{\text{tot}}^3 \rightarrow \infty$ as $N, Z_{\text{tot}} \rightarrow \infty$ with N/Z_{tot} fixed, then*

$$E^Q(N, B, Z_1, \dots, Z_K)/E^Q(N, B, Z_{\text{tot}}) \rightarrow 1. \quad (57)$$

The molecular binding energy is defined by

$$E_b^Q(N, B, Z_1, \dots, Z_K) = \min\{E^Q(N^{\{a\}}, B, Z^{\{a\}}) + E^Q(N^{\{b\}}, B, Z^{\{b\}}) - E^Q(N, B, Z_1, \dots, Z_K)\}. \quad (58)$$

The minimum is over decompositions of $\{1, \dots, K\}$ into two clusters, $\{a\}$ and $\{b\}$, and $Z^{\{a\}}$ and $Z^{\{b\}}$ stand for n -tuples of nuclear charges corresponding to each cluster, while $N^{\{a\}} + N^{\{b\}} = N$. For neutral molecules of identical atoms we obtain as an immediate corollary of Theorems 2.5 and 2.8:

Theorem 2.9 (Strong molecular binding) *The binding energy of a neutral molecule of K nuclei with charge Z satisfies*

$$\frac{E_b^Q(KZ, B, Z, \dots, Z)}{K|E^Q(Z, B, Z)|} \rightarrow 3[K/2](K - [K/2]) \quad (59)$$

as $Z \rightarrow \infty$ and $B/Z^3 \rightarrow \infty$. Here $[K/2]$ denotes the integer part of $K/2$.

2.6 Ingredients for the Proofs

The chain of arguments leading to the proofs of the quantum mechanical limit theorems in Sects. 2.3–5 is rather long and it is not possible here to mention but the main links. A first remark is that it suffices to prove the theorems about the *energy* asymptotics, but with slightly more general potentials than $-Z/|x|$, for the limit theorems for the density can then be obtained in a standard way by variation with respect to the potential [9]. To prove Theorems 2.1 and 2.3 one must derive upper and lower bounds on the quantum mechanical energy in terms of ground-state energies of the relevant functionals with controlled errors. The upper bounds are easier than the lower bounds and are obtained using the variational principle of Lieb [23]. This principle implies that for any kernel $\mathcal{K}(x, s; x', s')$ in the space and spin variables, satisfying $0 \leq \mathcal{K} \leq I$ as an operator on $L^2(\mathbf{R}^3; \mathbf{C}^2)$ as well as $\text{Tr } \mathcal{K} \leq N$, one has

$$E^Q(N, B, Z) \leq \text{Tr}[(H_A - Z/|x|^{-1})\mathcal{K}] + D(\rho_{\mathcal{K}}, \rho_{\mathcal{K}}) \quad (60)$$

where $\rho_{\mathcal{K}}(x) = \sum_s \mathcal{K}(x, s; x, s)$. Suitable kernels $\mathcal{K}$ are in the semiclassical Regions 1–3 constructed by means of *coherent states*, which may be regarded as a rigorous version of the “private rooms” used in the heuristic arguments in Sect. 2.2. Coherent states are usually understood as maps from the classical phase space of a system into wave functions, or rather into rank one projectors, having optimal localization around the classical values. The objects used here are slightly more general and are reminiscent of the coherent operators considered in [24]. They are defined by the integral kernels

$$\Pi_{\nu, u, p}(x, s; x', s') = g_r(x - u) \Pi_{\nu}^{\perp}(x_{\perp}, x'_{\perp}) e^{ip(x_3 - x'_3)} g_r(x' - u) \delta_{ss'}. \quad (61)$$

Here ν is a Landau level index, $u \in \mathbf{R}^3$, $p \in \mathbf{R}$, $\Pi_\nu^\perp(x_\perp, x'_\perp)$ is the integral kernel of the projector on the ν th Landau level in $L^2(\mathbf{R}^2, dx_\perp)$, and $g_r(x) = r^{-3/2}g(x/r)$ with $\int g^2 = 1$ is a localization function. These operators were first used in [17] to prove Theorem 2.1 for the special case of Regions 1 and 2. They are positive, although not projectors, and satisfy the coherent operator identities

$$(2\pi)^{-1} \sum_\nu \int \int \Pi_{\nu,u,p} du dp = I \quad \text{and} \quad \text{Tr}(\Pi_{\nu,u,p}) = d_\nu(B). \quad (62)$$

Moreover, they are approximate eigenoperators for both H_A and local potentials V in the sense that

$$\text{Tr}(H_A \Pi_{\nu,u,p}) = d_\nu(B) \varepsilon_{\nu,p} + d_\nu(B) r^{-2} \int (\nabla g)^2, \quad (63)$$

where $\varepsilon_{\nu,p} = 2\nu B + p^2$ is a (generalized) eigenvalue of H_A , and

$$\text{Tr}(V \Pi_{\nu,u,p}) = d_\nu(B) V * g_r^2. \quad (64)$$

The test kernel $\mathcal{K}$ is defined by

$$\mathcal{K}(x, s; x', s') = \delta_{ss'} \sum_\nu \int \int \Pi_{\nu,u,p}(x, s; x', s') \Theta(\tau'_B(\rho^{\text{MTF}}(u)) - \varepsilon_{\nu,p}) du dp, \quad (65)$$

where Θ is the Heaviside function, $\Theta(t) = 1$ for $t \geq 0$ and 0 for $t < 0$. Using (58), (59) and the properties of ρ^{MTF} it is easy to show that if the localization length r is chosen $\sim (Z^{-1/3}(1+\beta)^{-2/5})^\gamma$ with $0 < \gamma < 1$ the errors in the upper bound (55) are small compared to the main contribution $E^{\text{MTF}}(N, B, Z)$.

For the upper bound in Regions 4 and 5 one cannot use the coherent states (56) because the localization error due to the function g becomes too large. Instead one makes use of the fact that the wave functions in the lowest Landau level are themselves localized on a length scale $\sim \ell_B$, which, by the heuristic arguments of Sect. 2.2, is small compared with the atomic dimensions perpendicular to the field. The kernel $\mathcal{K}$ is here defined as

$$\mathcal{K}(x, s; x', s') = \delta_{ss'} d_0(B)^{-1} \int \Pi_0^\perp(x_\perp, y_\perp) \Gamma_{y_\perp}^{\text{DM}}(x_3, x'_3) \Pi_0^\perp(y_\perp, x'_\perp) dy_\perp. \quad (66)$$

To prove that the error is of lower order than the ground-state energy one uses an independent simple upper bound for E^Q in terms of the energy without electronic repulsion but with a reduced nuclear charge.

The lower bounds on the quantum mechanical energy require more tools. In Regions 1–3 one uses coherent states in a similar way as in the upper bound and, as an essential new ingredient, a generalization of the Lieb–Thirring estimate on the sum of negative eigenvalues of Schrödinger Hamiltonians [25].

Theorem 2.10 (Magnetic LT-inequality) *Let $|V|_- \in L^{5/2}(\mathbf{R}^3) \cap L^{3/2}(\mathbf{R}^3)$, where $|V(x)|_- = |V(x)|$ for $V \leq 0$ and 0 otherwise, and let $e_j(B, V)$, $j = 1, 2, \dots$ denote the negative eigenvalues of the operator $H = H_A + V(x)$. Then*

$$\sum_j |e_j(B, V)| \leq L_1 B \int |V(x)|_-^{3/2} dx + L_2 \int |V(x)|_-^{5/2} dx, \quad (67)$$

for certain constants L_1 and L_2 .

For generalizations to inhomogeneous field and a discussion of the constants in (62) see [26]. The Lieb–Thirring inequality (62) implies via Legendre transformation an estimate of the kinetic energy $T_\Psi = (\Psi, \sum_i H_A^{(i)} \Psi)$ from below by $(\text{const.}) \int \tau_B(\rho_\Psi)$. These estimates are needed both to control the localization error, when the Coulomb singularity is smeared as in (59) by g_r^2 , and the indirect (i.e., exchange-correlation) part of the Coulomb repulsion, which, by the Lieb–Oxford inequality [27] is bounded below by $-(\text{const.}) \int \rho_\Psi^{4/3}$.

In Regions 4 and 5 there are two main ingredients for the lower bound. The first is a theorem on the confinement of the electrons in the lowest Landau band as $B/Z^{4/3} \rightarrow \infty$.

Theorem 2.11 (Confinement to lowest Landau band) *Define*

$$E_{\text{conf}}^Q(N, Z, B) = \inf_{(\Psi, \Psi)=1} (\Psi, \Pi_0^N H_{N,B,Z} \Pi_0^N \Psi) \quad (68)$$

where Π_0^N is the projector on the lowest Landau band for the N -electron system. Then

$$E^Q/E_{\text{conf}}^Q \rightarrow 1 \quad (69)$$

as $B/Z^{4/3} \rightarrow \infty$ with N/Z fixed.

The second ingredient is an estimate of the indirect part of the Coulomb repulsion for strong magnetic fields. The Lieb–Oxford inequality, although universally true, is not strong enough for $B \sim Z^5$ and larger. In fact, if ρ is essentially concentrated in a cylinder of radius $\sim (Z/B)^{1/2}$ and length $\sim [Z \ln \eta]^{-1}$, as ρ_Ψ will eventually be according to the heuristics in Sect. 2.2, then $\int \rho^{4/3}$ is of the order $Z^{4/3} B^{1/3} [\ln \eta]^{1/3}$. Only for $B \ll Z^5$ will this be smaller than the expected ground-state energy $\sim Z^3 [\ln \eta]^2$. The indirect part of the Coulomb energy is essentially N times the self-energy of a unit charge smeared over the private room of an electron and should accordingly be of the order $Z^2 \ln(B/Z^3) \ln(B/Z^2)$. The estimate derived in [3] does not quite reach this ideal. The bound is obtained by introducing two ultraviolet cut-offs in the Coulomb kernel, one for the directions perpendicular to the field and another in the direction of the field, and it involves the kinetic energy. The result is

Theorem 2.12 (Bound on the indirect Coulomb energy)

$$\begin{aligned} H_{N,B,Z} &\geq \sum_{i=1}^N \left((1 - Z^{-1/3}) H_A^{(i)} + V^{\text{DM}}(x^{(i)}) \right) - D(\rho^{\text{DM}}, \rho^{\text{DM}}) \\ &\quad - C_\lambda (1 + \lambda^5) (1 + Z^{8/3}) (1 + [\ln \eta]^2) \end{aligned}$$

where V^{DM} is the potential generated by the nucleus and ρ^{DM} , and C_λ is a constant depending only on $\lambda = N/Z$.

In addition to Theorems 2.7–9 good control over the asymptotic MTF and DM theories is important for the proofs of the limit theorems. In the MTF case the methods of [9] and [10] can be used, but the DM theory requires new ideas. In particular, establishing the existence and uniqueness of the minimizer Γ^{DM} is not entirely straightforward.

3 Quantum Dots

3.1 The Hamiltonian

A quantum dot is modeled as a two-dimensional system of N electrons with effective mass m_* in a continuous potential V with $V(x) \rightarrow \infty$ for $|x| \rightarrow \infty$. (For instance, $V(x) = K|x|^2$.) The two-dimensional medium has dielectric constant ϵ and the effective charge for the electronic Coulomb interactions is $e_* = e/\sqrt{\epsilon}$. A magnetic field of strength B points in the direction perpendicular to the two-dimensional plane. Units are chosen such that $\hbar = 2m_* = e_* = 1$, and B is measured in units of $4B_*$ with $B_* = e_*^3 m_*^2 c / (\epsilon^{1/2} \hbar^2)$. To simplify notation the electron spin will here be ignored. The Hamiltonian can then be written

$$H_{N,B,V} = \sum_{i=1}^N \left[\left(\mathbf{p}^{(i)} + \mathbf{A}(x^{(i)}) \right)^2 - B + V(x^{(i)}) \right] + \sum_{i < j} |x^{(i)} - x^{(j)}|^{-1} \quad (70)$$

where $\mathbf{p} = -i(\partial_1, \partial_2)$, $x = (x_1, x_2) \in \mathbf{R}^2$ and $\mathbf{A}(x) = \frac{1}{2}(-x_2 B, x_1 B)$ as before. The spectrum of the kinetic energy operator $H_{\text{kin}} = (\mathbf{p} + \mathbf{A})^2 - B$ consists of the Landau levels $2\nu B$, $\nu = 0, 1, \dots$, where B has been subtracted for convenience in order that the spectrum of H_{kin} starts at 0. We write the potential as $V(x) = Kv(x)$ with a coupling constant K , and regard v as fixed, while K varies proportionally to N . The ground-state energy and electronic density for the Hamiltonian (65) are denoted respectively by $E^{\text{Q}}(N, B, K)$ and $\rho_{N,B,K}^{\text{Q}}(x)$.

3.2 The Asymptotic Theories

The main features of the asymptotics of $E^{\text{Q}}(N, B, K)$ and $\rho_{N,B,K}^{\text{Q}}(x)$ can be made plausible by simple heuristic arguments in the spirit of Sect. 2.2. Here, however, it is necessary to include the Coulomb repulsion between the electrons to get the right picture of the size of the dot for strong B . The reason is twofold: First, the kinetic energy vanishes in the lowest Landau level since there is no degree of freedom along the field as in the three dimensional case. Second, the potential V is nonsingular by assumption and the repulsion is therefore not compensated by an attractive singularity as for natural atoms.

Starting at $B = 0$, the kinetic energy of an electron is $\varepsilon_{\text{kin}} \sim N/R^2$, with R the radius of the dot, the potential energy from the confining V is $\varepsilon_{\text{pot}} \sim Kv(R)$ and the repulsive energy $\varepsilon_{\text{rep}} \sim N/R$. If $K \rightarrow \infty$ with $K/N = k$ fixed as $N \rightarrow \infty$, all these terms are proportional to N and the optimal radius is thus to leading order *independent of N* . (It depends on k though, and tends to infinity if $k \rightarrow 0$.) The energy of an electron is $\sim N$ and the total ground-state energy is thus $E \sim N^2$. The division line between weak and strong fields is clearly $B \sim N$: for $B \ll N$ the energy differences between Landau levels are much smaller than the other energies, whereas for $B \gg N$ the electrons will essentially sit in the lowest Landau level and the kinetic energy will be close to zero. An interesting point is that the electronic repulsion and the absence of attractive singularities of V prevents the dot from shrinking indefinitely as $B/N \rightarrow \infty$, contrary to what happens if $B/Z^3 \rightarrow \infty$ for natural, 3D atoms. In fact, dropping the kinetic term in (65) altogether leads to a model of N classical point particles interacting by Coulomb repulsion and with V . The radius of this system is approximately the value of R that minimizes $kv(R) + 1/R$, and while it is slightly smaller than the $B = 0$ value, where the kinetic term $1/R^2$ pushes the radius up, it is independent of B . Note also that for $B \gg N$ the electrons are localized on the scale $\ell_B \sim B^{-1/2}$ which is much smaller than the mean electronic distance $\sim N^{-1/2}$, while in the 3D case, with the potential $-Z/|x|$, the hard core bound $R \geq N^{1/2}\ell_B$ is saturated in the asymptotic ground state.

The density functionals, that enable to turn the heuristic reasoning above into precise theorems about the $N, K \rightarrow \infty$ limit, are defined as follows. For weak fields, $B \ll N$, a standard two-dimensional Thomas–Fermi energy functional applies,

$$\mathcal{E}^{\text{TF}}[\rho; K] = 2\pi \int \rho(x)^2 dx + \int V(x)\rho(x)dx + D(\rho, \rho). \quad (71)$$

At moderate fields, $B \sim N$, the correct asymptotics is given by a two-dimensional magnetic TF functional

$$\mathcal{E}^{\text{MTF}}[\rho; B, K] = \int j_B(\rho(x))dx + \int V(x)\rho(x)dx + D(\rho, \rho) \quad (72)$$

where j_B is a piece-wise linear function, defined by $j'_B(\rho) = [2\pi\rho/B]$ where $[t]$ denotes the integral part of t . This functional was first applied to quantum dots in [28] and is capable of modeling interesting details in the conductivity of large quantum dots in magnetic fields [29]. The last asymptotic density functional, the “classical continuum model”, applies to strong fields, $B \gg N$. It contains only classical interactions of the charge density and no kinetic energy term:

$$\mathcal{E}^{\text{C}}[\rho; K] = \int V(x)\rho(x)dx + D(\rho, \rho). \quad (73)$$

In addition to the three density functionals the model of N point charges with energy

$$\mathcal{E}^{\text{P}}(x_1, \dots, x_N) = \sum_i V(x_i) + \sum_{i < j} |x_i - x_j|^{-1} \quad (74)$$

is also relevant. The ground-state energies of (66)–(68), i.e. the infima (in fact minima) of the functionals for densities with given $\int \rho = N$, are denoted $E^{\text{TF}}(N, K)$, $E^{\text{MTF}}(N, B, K)$ and $E^{\text{C}}(N, K)$, and an analogous notation is used for the minimizing densities. The infimum of (69) for given N is denoted $E^{\text{P}}(N, K)$. The energies and densities scale as expected from the heuristic arguments, e.g., $E^{\text{TF}}(N, K) = N^2 E^{\text{TF}}(1, K/N)$, $E^{\text{C}}(N, K) = N^2 E^{\text{C}}(1, K/N)$, $E^{\text{MTF}}(N, B, K) = N^2 E^{\text{MTF}}(1, B/N, K/N)$, $\rho_{N,B,K}^{\text{MTF}}(x) = N \rho_{1,B/N,K/N}^{\text{MTF}}(x)$ etc.

The difference between $E^{\text{C}}(N, K)$ and $E^{\text{P}}(N, K)$ can be estimated, using an electrostatic lemma of Lieb and Yau [30], as

$$E^{\text{C}}(N, K) - aN^{3/2} \geq E^{\text{P}}(N, K) \geq E^{\text{C}}(N, K) - bN^{3/2}, \quad (75)$$

where $a, b > 0$ depend only on $k = K/N$. Note that the order $N^{3/2} = N \times (N^{-1/2})^{-1}$ is precisely what is expected from the indirect part of the Coulomb interaction (which is absent in E^{C}).

In order to guarantee that the minimizer of (68) is an L^1 function and not just a measure, V is taken to be in the class $C^{1,\alpha}$ of once differentiable functions whose derivatives are Lipschitz continuous of some order $\alpha > 0$. For the important special case $V(x) = K|x|^2$ the minimizer $\rho_{N,K}^{\text{C}}$ can be computed explicitly; it has the shape of a half-ellipsoid:

$$\rho_{N,V}^{\text{C}}(x) = \frac{3}{2\pi} N R^{-2} \sqrt{1 - R^{-2}|x|^2} \quad (76)$$

if $|x| \leq R = (3\pi N/8K)^{1/3}$ and 0 otherwise.

The shape of the electronic density in the three asymptotic theories is shown in Fig. 3 for a quadratic potential $V(x) = K|x|^2$ and different values of B . The pictures were prepared by K. Johnsen. At $B = 0$ we have the usual Thomas–Fermi model, which may be regarded as a limit of the MTF model with infinitely many Landau levels occupied. As the field increases the number of occupied Landau levels decreases and the full levels show up as plateaus at integer values of $2\pi\rho/B$. At $B = 2$ T three Landau levels are full and electrons in the central dome occupy the fourth level. At $B = 4$ T most electrons are in the lowest level, but some are in the next highest level. At $B = 8$ T, the density is a half ellipsoid, given by the minimizer (71) of the classical functional (68). Note that the diameter of the dot at $B = 8$ T is slightly smaller than at $B = 0$, but as soon as the region of the model (68) has been reached the density does not change further although B is increased.

3.3 The Limit Theorem

The analogue for quantum dots of the limit theorems in Sects. 2.3–5 is

Theorem 3.1 (Limit theorem for the energy and density) *Let $V = Kv$ with v a fixed function in $C^{1,\alpha}$, $0 < \alpha$. Then, if $N, K \rightarrow \infty$ with $K/N = k$ fixed,*

$$E^{\text{Q}}(N, B, K)/E^{\text{MTF}}(N, B, K) \rightarrow 1 \quad (77)$$

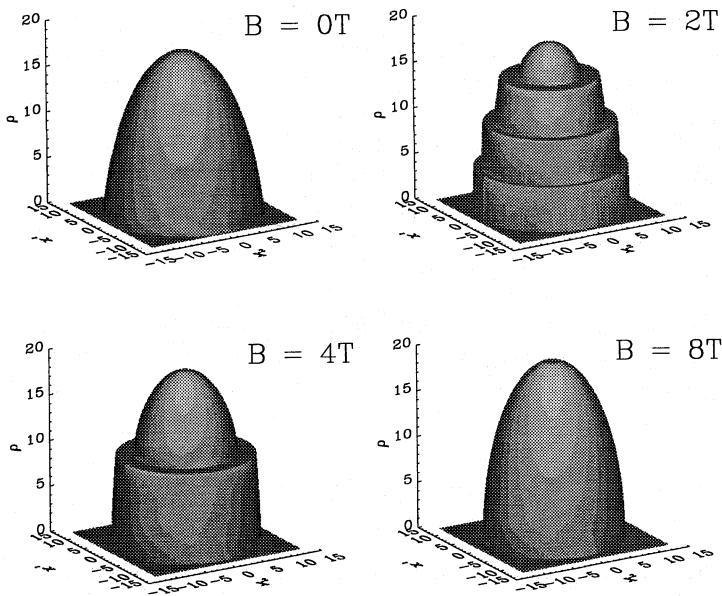


Figure 3: Quantum dots at various magnetic field strengths. The potential is $V(x) = (1/2)m_*\omega^2|x|^2$, with $m_* = 0.67m$, $\hbar\omega = 3.37$ meV, and $N = 50$. The coordinate axes are displayed in units of 10^{-8} m and the density ρ in the units 10^{-14} m $^{-2}$.

uniformly in B . Moreover,

$$E^Q(N, B, K)/E^{\text{TF}}(N, K) \rightarrow 1 \quad \text{if } B/N \rightarrow 0 \quad (78)$$

and

$$E^Q(N, B, K)/E^{\text{C}}(N, K) \rightarrow 1 \quad \text{if } B/N \rightarrow \infty. \quad (79)$$

The densities converge also:

$$\frac{1}{N}\rho_{N,B,K}^Q \rightarrow \rho_{1,B/N,k}^{\text{MTF}} \quad (80)$$

uniformly in B , and

$$\frac{1}{N}\rho_{N,B,K}^Q \rightarrow \rho_{1,k}^{\text{TF}} \quad \text{if } B/N \rightarrow 0 \quad (81)$$

$$\frac{1}{N}\rho_{N,B,K}^Q \rightarrow \rho_{1,k}^{\text{C}} \quad \text{if } B/N \rightarrow \infty. \quad (82)$$

The convergence is in the weak L^1 sense.

The proof of this theorem proceeds partly along similar lines as in the 3D case. Thus the upper bound on the energy is obtained using coherent states as in (56) (without the $\exp ip(x_3 - x'_3)$ factor, of course). For the lower bound one treats the cases $B \ll N$ and $B \sim N$ differently from $B \gg N$. For strong fields the Hamiltonian is estimated from below by the point charge functional (69), which in turn is bounded by the classical functional with a small error according to (70). For weak and moderate fields one uses coherent states, and the indirect Coulomb energy is estimated by $-(\text{const.}) \int \rho^{3/2}$. The following inequality is needed to control this term.

Theorem 3.2 (2D magnetic Lieb–Thirring inequality) *Let W be locally integrable, and denote by $e_1(W), e_2(W), \dots$ the negative eigenvalues (if any) of the operator $H = (\mathbf{p} + \mathbf{A})^2 - B + W$. Then for all $0 < \lambda < 1$ we have the estimate*

$$\sum_j |e_j(W)| \leq \lambda^{-1} \frac{B}{4\pi} \int_{\mathbf{R}^2} |W(x)|_- dx + \frac{3}{8} (1 - \lambda)^{-2} \int_{\mathbf{R}^2} |W(x)|_-^2 dx. \quad (83)$$

The proof is slightly more complicated than in the 3D case because the quantum mechanical kinetic energy vanishes in the lowest Landau level.

Acknowledgments. Thanks are due to K. Johnsen and Ö. Rögnvaldsson for the pictures presented here, and to I. Fushiki, E. H. Gudmundsson, V. Gudmundsson, J. Kinaret, M. Loss and C. Pethick for valuable discussions. This work was supported by US NSF grants Nos. PHY90-19433 (EHL) and DMS 92-03829 (JPS), and by the Icelandic Science Foundation (JY).

References

- [1] G. Chanmugam, *Magnetic Fields of Degenerate Stars*, Ann. Rev. Astron. Astrophys. **30**, 143–184 (1992)
- [2] E.H. Lieb, J.P. Solovej and J. Yngvason, *Asymptotics of Heavy Atoms in High Magnetic Fields: I. Lowest Landau Band Regions*, Commun. Pure Appl. Math. **47**, 513–591 (1994)
- [3] E.H. Lieb, J.P. Solovej and J. Yngvason, *Asymptotics of Heavy Atoms in High Magnetic Fields: II. Semiclassical Regions*, Commun. Math. Phys **161**, 77–124 (1994)
- [4] M.A. Kastner, *Artificial atoms*, Phys. Today **46**, 24–31 (1993)
- [5] E.H. Lieb, J.P. Solovej and J. Yngvason, *The Ground States of Large Quantum Dots in Magnetic Fields*, Phys. Rev. B **51**, 10646–10665 (1995)
- [6] E.H. Lieb, J.P. Solovej and J. Yngvason, *Heavy Atoms in the Strong Magnetic Field of a Neutron Star*, Phys. Rev. Lett. **69**, 749–752 (1992)

- [7] E.H. Lieb and J.P. Solovej, *Atoms in the Magnetic Field of a Neutron Star*, in: *Differential Equations with Applications to Mathematical Physics*, W.F. Ames, J.V. Herod and E.M. Harrell II, eds., pp. 221–237, Academic Press 1993
- [8] E.H. Lieb, J.P. Solovej and J. Yngvason, *Quantum Dots*, in: *Proceedings of the Conference on Partial Differential Equations and Mathematical Physics*, University of Alabama, Birmingham, 1994, I. Knowles, ed., pp. 157–172, International Press 1995
- [9] E.H. Lieb and B. Simon, *The Thomas–Fermi Theory of Atoms, Molecules and Solids*, *Adv. in Math.* **23**, 22–116 (1977)
- [10] E.H. Lieb, *Thomas–Fermi and related theories of atoms and molecules*, *Rev. Mod. Phys.* **53**, 603–641 (1981); *Erratum*, *Rev. Mod. Phys.* **54**, 311 (1982)
- [11] B.B. Kadomtsev, *Heavy Atoms in an Ultrastrong Magnetic Field*, *Soviet Phys. JETP* **31**, 945–947 (1970)
- [12] B.B. Kadomtsev and V.S. Kudryavtsev, *Atoms in a superstrong magnetic field*, *JETP Lett.* **13**, 42–44 (1971)
- [13] M. Ruderman, *Matter in Superstrong Magnetic Fields: The Surface of a Neutron Star*, *Phys. Rev. Lett.* **27**, 1306–1308 (1971).
- [14] R.O. Mueller, A.R.P. Rau and L. Spruch, *Statistical Model of Atoms in Intense Magnetic Fields*, *Phys. Rev. Lett.* **26**, 1136–1139 (1971)
- [15] Y. Tomishima and K. Yonei, *Thomas–Fermi Theory for Atoms in a Strong Magnetic Field*, *Progr. Theor. Phys.* **59**, 683–696 (1978)
- [16] B. Banerjee, D.H. Constantinescu, and P. Rehák, *Thomas–Fermi and Thomas–Fermi–Dirac calculations for atoms in a very strong magnetic field*, *Phys. Rev. D* **10**, 2384–2395 (1974)
- [17] J. Yngvason, *Thomas–Fermi Theory for Matter in a Magnetic Field as a Limit of Quantum Mechanics*, *Lett. Math. Phys.* **22**, 107–117 (1991)
- [18] V. Ivrii, *Semiclassical Microlocal Analysis and Spectral Asymptotics*, Springer (to be published)
- [19] A. Sobolev, *The quasi-classical asymptotics of local Riesz means for the Schrödinger operator in a strong homogeneous magnetic field*, *Duke Math. J.* **74**, 319–429 (1994)
- [20] I. Fushiki, E.H. Gudmundsson, C.J. Pethick, and J. Yngvason, *Matter in a Magnetic Field in the Thomas–Fermi and Related Theories*, *Ann. Phys.* **216**, 29–72 (1992)

- [21] Ö.E. Rögnvaldsson, I. Fushiki, C.J. Pethick, E.H. Gudmundsson and J. Yngvason, *Thomas–Fermi Calculations of Atoms and Matter in Magnetic Neutron Stars: Effects of Higher Landau Bands*, *Astrophys. J.* **216**, 276–290 (1993)
- [22] K. Johnsen, MS thesis, Univ. of Iceland, 1994. See also: K. Johnsen and J. Yngvason, *Density Matrix Functional Calculations for Matter in Strong Magnetic Fields: Ground States of Heavy Atoms*, *Phys. Rev. A*, in press (1996)
- [23] E.H. Lieb, *A Variational Principle for Many-Fermion Systems*, *Phys. Rev. Lett.* **46**, 457–459; Erratum **47**, 69 (1981)
- [24] E.H. Lieb, J.P. Solovej, *Quantum coherent operators: A generalization of coherent states*, *Lett. Math. Phys.* **22**, 145–154 (1991)
- [25] E.H. Lieb and W.E. Thirring, *Bound for the Kinetic Energy of Fermions Which Proves the Stability of Matter*, *Phys. Rev. Lett.* **35**, 687–689 (1975)
- [26] L. Erdős, *Magnetic Lieb–Thirring inequalities*, *Comm. Math. Phys.* **170**, 629–669 (1995)
- [27] E.H. Lieb and S. Oxford, *An Improved Lower Bound on the Indirect Coulomb Energy*, *Int. J. Quant. Chem.* **19**, 427–439 (1981)
- [28] P.L. McEuen, E.B. Foxman, J. Kinaret, U. Meirav, M.A. Kastner, N.S. Wingreen and S.J. Wind, *Self consistent addition spectrum of a Coulomb island in the quantum Hall regime*, *Phys. Rev. B* **45**, 11419–11422 (1992)
- [29] N.C. van der Vaart, M.P. de Ruyter van Steveninck, L.P. Kouwenhoven, A.T. Johnson, Y.V. Nazarov, and C.J.P.M. Harmans, *Time-Resolved Tunneling of Single Electrons between Landau Levels in a Quantum Dot*, *Phys. Rev. Lett.* **73**, 320–323 (1994)
- [30] E.H. Lieb and H.-T. Yau, *The stability and instability of relativistic matter*, *Commun. Math. Phys.* **118**, 177–213 (1988)

Ground states of large quantum dots in magnetic fields

Elliott H. Lieb and Jan Philip Solovej

Department of Mathematics, Fine Hall, Princeton University, Princeton, New Jersey 08544

Jakob Yngvason

Science Institute, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland

(Received 1 December 1994)

The quantum-mechanical ground state of a two-dimensional (2D) N -electron system in a confining potential $V(x) = Kv(x)$ (K is a coupling constant) and a homogeneous magnetic field B is studied in the high-density limit $N \rightarrow \infty$, $K \rightarrow \infty$ with K/N fixed. It is proved that the ground-state energy and electronic density can be computed *exactly* in this limit by minimizing simple functionals of the density. There are three such functionals depending on the way B/N varies as $N \rightarrow \infty$: A 2D Thomas-Fermi (TF) theory applies in the case $B/N \rightarrow 0$; if $B/N \rightarrow \text{const} \neq 0$ the correct limit theory is a modified B -dependent TF model, and the case $B/N \rightarrow \infty$ is described by a classical continuum electrostatic theory. For homogeneous potentials this last model describes also the weak-coupling limit $K/N \rightarrow 0$ for arbitrary B . Important steps in the proof are the derivation of a Lieb-Thirring inequality for the sum of eigenvalues of single-particle Hamiltonians in 2D with magnetic fields, and an estimation of the exchange-correlation energy. For this last estimate we study a model of classical point charges with electrostatic interactions that provides a lower bound for the true quantum-mechanical energy.

I. INTRODUCTION

In the past few years considerable experimental and theoretical work has been devoted to the study of quantum dots, which are atomlike two-dimensional systems, confined within semiconductor heterostructures. The number of articles on this subject is by now quite large. See, e.g., Refs. 1 and 2 for reviews, Refs. 3–7 for recent measurements of conductivity and capacity of quantum dots, and Refs. 8–18 for various theoretical aspects and further references. The parameters of such artificial atoms may differ appreciably from their natural counterparts because of the interactions of the electrons with the crystal where they reside. In a quantum dot the natural atomic unit of length is $a_* = \epsilon \hbar^2 / (m_* e^2)$, where ϵ is the dielectric constant and m_* is the effective electron mass. Compared with the usual Bohr radius $a_0 = \hbar^2 / (me^2)$, the length a_* is typically large, e.g., $a_* \approx 185a_0$ in GaAs. The corresponding natural unit B_* with which we measure the magnetic field B is the field at which the magnetic length $l_B = \hbar e / (B^{1/2} c)$ equals a_* , i.e., $B_* = (a_0/a_*)^2 B_0$, where $B_0 = e^2 m^2 c / \hbar^3 = 2.35 \times 10^5$ T is the value corresponding to free electrons. If a_0/a_* is small, B_* can be much smaller than B_0 . Thus $B_* \approx 7$ T in GaAs. This makes it possible to study in the laboratory effects which, for natural atoms, require the magnetic fields of white dwarfs or even neutron stars.

The ground-state properties of natural atoms in high magnetic fields have recently been analyzed rigorously in the asymptotic limit where the number of electrons and the nuclear charge are large.^{19–21} For artificial atoms one may expect asymptotic analysis to be even more useful because the accuracy increases with the number of electrons, and a quantum dot can easily accommodate several hundred or even a thousand electrons. In the present pa-

per we carry out such an analysis of the ground state of a quantum dot in a magnetic field. One of our conclusions is that the self-consistent model introduced by McEuen *et al.*^{3,13} is a rigorous limit of quantum mechanics. This model has recently been applied to explain interesting features of the addition spectra of large quantum dots in strong magnetic fields.^{15,6}

Before discussing our results for dots we summarize, for comparison, the main findings about atoms in Refs. 19–21. The quantum-mechanical ground-state energy and electronic density of a natural atom or ion with electron number N and nuclear charge Z in a homogeneous magnetic field B can, in the limit $N \rightarrow \infty$, $Z \rightarrow \infty$ with Z/N fixed, be described exactly by functionals of the density, or, in one case, of density matrices. There are five different functionals, depending on the way B varies with N as $N \rightarrow \infty$. In each of the cases $B \ll N^{4/3}$ (with B measured in the natural unit B_0), $B \sim N^{4/3}$, and $N^{4/3} \ll B \ll N^3$ the correct asymptotics is given by an appropriate functional of the semiclassical Thomas-Fermi type. For $B \sim N^3$ a modified functional, depending on density matrices, is required, whereas the case $B \gg N^3$ is described by a density functional that can be minimized in closed form.

A review of our results about quantum dots was given in Ref. 22. Due to the reduced dimensionality of the electronic motion, there are only three different asymptotic theories for quantum dots instead of five for natural atoms. These three theories are given by simple functionals of the density and correspond, respectively, to the cases $B \ll N$, $B \sim N$, and $B \gg N$ (B measured in units of B_*) as $N \rightarrow \infty$ with V/N fixed, where V is the attractive exterior potential that restricts the two-dimensional motion of the electrons. This potential, which plays the same role as the nuclear attraction in a natural atom, is

generated in a quantum dot by exterior gates, and thus is adjustable to a certain extent. In the course of proving the asymptotic limits we shall also consider, in addition to the density functionals, a model of classical point charges in two dimensions that gives a lower bound to the quantum-mechanical energy.

Some of the methods and results of the present paper contrast markedly with those of our earlier work.^{19–21} From a mathematical point of view the most interesting feature of quantum dots compared to natural atoms is the somewhat peculiar electrostatics that appear because the interaction between the electrons is given by the three-dimensional Coulomb potential although the motion is two dimensional. Also, the fact that the kinetic energy vanishes in the lowest Landau level requires additional mathematical effort in order to bound the kinetic energy from below by a functional of the density. We now describe in more detail the limit theorems to be proved in the sequel. A quantum dot with N electrons in a confining potential V and a homogeneous magnetic field B is modeled by the following Hamiltonian:

$$H_N = \sum_{j=1}^N H_1^{(j)} + \frac{e^2}{\epsilon} \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}, \quad (1.1)$$

with $x_i \in \mathbb{R}^2$ and where H_1 is the one-body Hamiltonian

$$H_1 = \frac{\hbar^2}{2m_*} \left[i \nabla - \frac{e}{\hbar c} \mathbf{A} \right]^2 + g_* \left[\frac{\hbar e}{2mc} \right] \cdot \mathbf{S} \cdot \mathbf{B} - \left[\frac{\hbar e}{2mc} \right] \left[\frac{m}{m_*} - \frac{|g_*|}{2} \right] B + V(x). \quad (1.2)$$

As before, e and m denote the charge and mass of a (free) electron, ϵ is the dielectric constant, m_* is the effective mass, and g_* is the effective g factor. The magnetic vector potential is $\mathbf{A}(x) = \frac{1}{2}(-Bx^2, Bx^1)$ [with $x = (x^1, x^2) \in \mathbb{R}^2$], $\mathbf{B} = (0, 0, B)$, and $\mathbf{S}$ is the vector of electron spin operators. The potential $V(x)$ is supposed to be continuous and confining, which is to say that $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$. It is not assumed to be circularly symmetric. The constant term in (1.2), $-\left[\frac{\hbar e}{2mc}\right]\left[(m/m_*) - |g_*|/2\right]B$, is included in order that the kinetic-energy operator $H_{\text{kin}} = H_1 - V(x)$ has a spectrum starting at zero. The Hilbert space is that appropriate for fermions with spin, the antisymmetric tensor product $\wedge^N L^2(\mathbb{R}^2; \mathbb{C}^2)$.

We define an effective charge by $e_* = e/\sqrt{\epsilon}$ and choose units such that $\hbar = m_* = e_* = 1$. The unit of length is then the effective Bohr radius $a_* = \hbar^2/(m_* e_*^2)$ and the unit of energy is $E_* = e_*^2/a_* = e_*^3 m_*/\hbar^2$. Moreover, the unit B_* for the magnetic field is determined by $\hbar e B_*/(m_* c) = E_*$, so $B_* = e_*^3 m_*^2 c/(\epsilon^{1/2} \hbar^3)$. The values for GaAs are $a_* = 9.8$ nm, $E_* = 12$ meV, and $B_* = 6.7$ T.

The true quantum-mechanical ground-state energy of H_N is denoted by $E^Q(N, B, V)$ and the true ground-state electron density by $\rho_{N,B,V}^Q(x)$. The density functionals that describe the asymptotics of E^Q and ρ^Q are of three types. The first is a standard two-dimensional Thomas-Fermi energy functional

$$\mathcal{E}^{\text{TF}}[\rho; V] = (\pi/2) \int \rho(x)^2 dx + \int V(x) \rho(x) dx + D(\rho, \rho) \quad (1.3)$$

with

$$D(\rho, \rho) = \frac{1}{2} \int \frac{\rho(x) \rho(y)}{|x - y|} dx dy. \quad (1.4)$$

Here ρ is a non-negative density on $\mathbb{R}^2$ and all integrals are over $\mathbb{R}^2$ unless otherwise stated. The second functional is a two-dimensional magnetic Thomas-Fermi functional

$$\mathcal{E}^{\text{MTF}}[\rho; B, V] = \int j_B[\rho(x)] dx + \int V(x) \rho(x) dx + D(\rho, \rho), \quad (1.5)$$

where j_B is a piecewise linear function that will be defined precisely in the next section. This functional is the two-dimensional analog of the three-dimensional magnetic Thomas-Fermi functional that was introduced in Ref. 23 and further studied in Refs. 24, 25, and 21. The present two-dimensional version was first stated in Ref. 3; these authors call it the self-consistent (SC) model. The repulsion term considered in Ref. 3 is slightly different from $D(\rho, \rho)$, since it has cutoffs at long and short distances. It is still positive definite as a kernel and our methods can easily be adapted to prove Theorems 1.1 and 1.2 with such cutoff Coulomb kernels.

The last asymptotic functional will be called the classical functional, since the kinetic-energy term is absent and only classical interactions remain:

$$\mathcal{E}^C[\rho; V] = \int V(x) \rho(x) dx + D(\rho, \rho). \quad (1.6)$$

The functionals (1.3) and (1.6) are in fact limiting cases of (1.5) for $B \rightarrow 0$ and $B \rightarrow \infty$, respectively. As discussed in detail later, for each functional there is a unique density that minimizes it under the constraint $\int \rho = N$. We denote these densities, respectively, by $\rho_{N,V}^{\text{TF}}(x)$, $\rho_{N,B,V}^{\text{MTF}}(x)$, $\rho_{N,V}^C(x)$, and the corresponding minimal energies by $E^{\text{TF}}(N, V)$, $E^{\text{MTF}}(N, B, V)$, and $E^C(N, V)$.

In order to relate E^Q to these other energies we take a high-density limit. This is achieved by letting N tend to infinity (which is a reasonable thing to do physically, since N can be several hundred) and we let V tend to infinity. The latter statement means that we fix a potential v and set $V = Nv$. With this understanding of $N, V \rightarrow \infty$ our main results are summarized in the following two theorems. [In order to prove these theorems we need to assume that V is sufficiently regular. The technical requirement is that V belongs to the class $C_{\text{loc}}^{1,\alpha}$ (see Theorem 3.2 for the definition of $C_{\text{loc}}^{1,\alpha}$).

1.1 THEOREM (limit theorem for the energy). Let $V = Nv$ with v a fixed function in $C_{\text{loc}}^{1,\alpha}$. Then

$$\lim_{N \rightarrow \infty} E^Q(N, B, V)/E^{\text{MTF}}(N, B, V) = 1 \quad (1.7)$$

uniformly in B . Moreover,

$$\lim_{N \rightarrow \infty} E^Q(N, B, V)/E^{\text{TF}}(N, V) = 1 \quad \text{if } B/N \rightarrow 0 \quad (1.8)$$

and

$$\lim_{N \rightarrow \infty} E^Q(N, B, V) / E^C(N, V) = 1 \text{ if } B/N \rightarrow \infty. \quad (1.9)$$

1.2 THEOREM (limit theorem for the density). Let $V = Nv$ with v a fixed function in $C_{loc}^{1,\alpha}$. Then

$$\frac{1}{N} \rho_{N,B,V}^Q \rightarrow \rho_{1,B/N,v}^{\text{MTF}} \quad (1.10)$$

uniformly in B , and

$$\frac{1}{N} \rho_{N,B,V}^Q \rightarrow \rho_{1,v}^{\text{TF}} \text{ if } B/N \rightarrow 0, \quad (1.11)$$

$$\frac{1}{N} \rho_{N,B,V}^Q \rightarrow \rho_{1,v}^C \text{ if } B/N \rightarrow \infty. \quad (1.12)$$

The convergence is in the weak L^1 sense. [By definition, a sequence of functions f_n converges to a function f in weak L^1 sense if $\int f_n g \rightarrow \int f g$ for all bounded (measurable) functions g .]

Let us add a few comments on these results. As discussed in the Sec. II, the energy E^{MTF} has the scaling property

$$E^{\text{MTF}}(N, B, V) = N^2 E^{\text{MTF}}(1, B/N, V/N). \quad (1.13)$$

Thus (1.7) is equivalent to

$$E^Q(N, V, B) = N^2 E^{\text{MTF}}(1, B/N, V/N) + o(N^2), \quad (1.14)$$

where the error term is uniformly bounded in B for V/N fixed. One expects the error to be $O(N^{3/2})$, which is the order of the exchange contribution to the Coulomb interaction, but our methods do not quite allow us to prove this. We do, however, show that for B/N larger than a critical value (depending on V/N) one has $E^{\text{MTF}}(N, B, V) = E^C(N, V)$ and

$$E^Q(N, V, B) \geq N^2 E^C(1, V/N) - bN^{3/2}, \quad (1.15)$$

where the coefficient b depends only on V/N .

The condition that V/N is fixed as $N \rightarrow \infty$ guarantees that the diameter of the electronic density distribution stays bounded as $N \rightarrow \infty$; thus the limit we are considering is really a high-density limit rather than simply a large- N limit. On the other hand, for a homogeneous potential V (e.g., quadratic, as is often assumed) one obtains also a nontrivial $N \rightarrow \infty$ limit for V fixed, if the lengths are suitably scaled. In fact, this limit is given by the classical functional (1.6). Intuitively this is easy to understand, for if an increase in N is not compensated by an increase in V the charge density spreads out and the kinetic-energy terms in (1.5) and (1.3) become negligible compared with the other terms. (The result again requires V to be in $C_{loc}^{1,\alpha}$.)

1.3 THEOREM (energy limit with a homogeneous potential). Assume that v is homogeneous of degree $s \geq 1$, i.e.,

$$v(\lambda x) = \lambda^s v(x).$$

Then

$$\lim_{N \rightarrow \infty} E^Q(N, B, Kv) / E^{\text{MTF}}(N, B, Kv) = 1 \quad (1.16)$$

uniformly in B and in K as long as K/N is bounded above. Moreover, if $K/N \rightarrow 0$ as $N \rightarrow \infty$, then

$$\lim_{N \rightarrow \infty} E^Q(N, B, Kv) / E^C(N, B) = 1 \quad (1.17)$$

uniformly in B .

One can also prove a limit theorem for the density in the case of homogeneous potentials. Since the formulation of such a theorem becomes somewhat complicated we refrain from doing this, but refer to Eqs. (2.14)–(2.16) below for the scaling of the MTF functional with $k = K/N$ and to (3.24) for the weak-coupling limit of the MTF density.

The proof of the limit theorems involves the following steps. In Secs. II and III we discuss the basic properties of the functionals (1.3)–(1.6). In Sec. IV we consider the energy of a system of classical point charged particles in $\mathbf{R}^2$ in the exterior potential V as a function of the positions of the charges. This energy has a minimum, denoted by $E^P(N, V)$ (with P denoting “particle”). A significant remark is that the charge configuration, for which the minimum is obtained, is confined within a radius independent of the total charge N for fixed V/N . This finite-radius lemma, which also holds for the charge densities minimizing the functionals (1.3)–(1.6), is proved in the Appendix. Using this and an electrostatics lemma of Lieb and Yau²⁶ we derive the bounds

$$E^C(N, V) - aN^{3/2} \geq E^P(N, V) \geq E^C(N, V) - bN^{3/2}, \quad (1.18)$$

where a and b depend only on V/N . These bounds are of independent interest apart from their role in the proof of the limit theorems where, in fact, only the latter inequality is needed. Upper and lower bounds to the quantum-mechanical energy $E^Q(N, B, V)$ in terms of $E^{\text{MTF}}(N, B, V)$ with controlled errors are derived in Sec. V. The upper bound is a straightforward variational calculation using magnetic coherent states in the same way as in Ref. 21. For the lower bound one treats the cases of large B and small B separately. The estimate for large B is obtained by first noting that obviously $E^Q(N, B, V) \geq E^P(N, V)$, because the kinetic energy is non-negative, and then using (1.18). For small B two auxiliary results are required: a generalization of the magnetic Lieb-Thirring equality considered in Ref. 21, and an estimate of the correlation energy. Once these have been established the proof of Theorem 1.1 is completed by a coherent state analysis. The limit theorem for the density follows easily from the limit theorem for the energy by perturbing V with bounded functions.

II. THE MTF THEORY: ITS DEFINITION AND PROPERTIES

By employing the natural units defined in the Introduction the kinetic-energy operator can be written

$$H_{\text{kin}} = \frac{1}{2} (i \nabla - \mathbf{A})^2 + \gamma \mathbf{S} \cdot \mathbf{B} - \frac{1}{2} (1 - |\gamma|) B \quad (2.1)$$

with $\gamma = g_* m_*/(2m)$. The spectrum of H_{kin} is

$$\epsilon_{n,\sigma} = (n + \gamma\sigma + \frac{1}{2} |\gamma|) B \quad (2.2)$$

with $n = 0, 1, \dots, \sigma = \pm \frac{1}{2}$. We write the energy levels (2.2) in strictly increasing order as $\epsilon_v(B)$, $v = 0, 1, \dots$. The degeneracy of each level per unit area is

$d_v(B) = B/(2\pi)$, except if, by coincidence, γ happens to be an integer; in that case $d_v(B) = B/(2\pi)$ for $v=0, \dots, |\gamma|-1$, while $d_v(B)$ is twice as large for the higher levels. It is worth recalling that if $V(x) = K|x|^2$, the spectrum of the one-body Hamiltonian H_1 in (1.2) is solvable. The spectrum of H_1 was determined by Fock²⁷ in 1928, two years before Landau's paper on the spectrum of $(i\nabla - \mathbf{A})^2$. For the Hamiltonian without spin, namely, $\frac{1}{2}(i\nabla - \mathbf{A})^2 + K|x|^2$, the spectrum is given by

$$E = \frac{1}{2}(n_1 - n_2)B + \frac{1}{2}(n_1 + n_2 + 1)[4K + B^2]^{1/2}$$

with $n_1, n_2 = 0, 1, 2, \dots$. It is remarkable that this simple spectrum gives a qualitatively good fit to some of the data.⁸

For a gas of noninteracting fermions with the energy spectrum (2.2) the energy density j_B as a function of the particle density ρ is given by

$$j_B(0) = 0,$$

$$j'_B(\rho) = \epsilon_v(B) \text{ if } D_v(B) < \rho < D_{v+1}(B), \quad v=0, 1, \dots,$$

where $j'_B = dj_B/d\rho$ and

$$D_v(B) = \sum_{\nu=0}^v d_\nu(B).$$

More explicitly,

$$j_B(\rho) = \sum_{\nu=0}^{\nu_{\max}} \epsilon_\nu(B) d_\nu(B) + [\rho - D_{\nu_{\max}}(B)] \epsilon_{\nu_{\max}+1}(B), \quad (2.3)$$

where $\nu_{\max} = \nu_{\max}(\rho, B)$ is defined by

$$D_{\nu_{\max}}(B) \leq \rho < D_{\nu_{\max}+1}(B).$$

Thus j_B is a convex, piecewise linear function with $j_B(\rho) = 0$ for $0 \leq \rho \leq d_1(B)$. It has the scaling property

$$j_B(\rho) = B^2 j_1(\rho/B). \quad (2.4)$$

As $B \rightarrow 0$, j_B becomes a quadratic function of the density:

$$\lim_{B \rightarrow 0} j_B(\rho) = j_0(\rho) = \frac{\pi}{2} \rho^2. \quad (2.5)$$

Moreover,

$$j_B(\rho) \leq j_0(\rho) \quad (2.6)$$

for all ρ and B (see Fig. 1).

Given an exterior potential V the MTF functional is defined by (1.5). We assume that V is continuous (V measurable and locally bounded would suffice) and tends to ∞ as $|x| \rightarrow \infty$. In particular, V is bounded below and, by adding a constant if necessary, we may assume that $V(x) \geq 0$ everywhere. Because of (2.6) the functional (1.5) is defined for all non-negative functions ρ such that $\int \rho V < \infty$, $\int \rho^2 < \infty$, and $D(\rho, \rho) < \infty$. Since $V \geq 0$ the functional is non-negative. If N is some positive number we denote

$$\mathcal{C}_N = \left\{ \rho : \rho \geq 0, \int \rho V < \infty, \int \rho^2 < \infty, D(\rho, \rho) < \infty, \int \rho = N \right\}$$

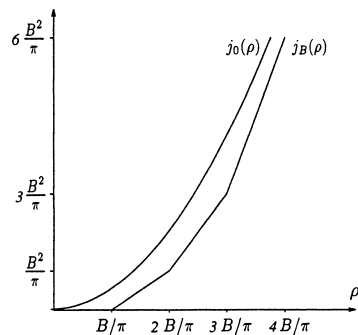


FIG. 1. The kinetic-energy densities $j_B(\rho)$ and $j_0(\rho)$ in the special case where $\gamma = 0$.

and define the MTF energy by

$$E^{\text{MTF}}(N, B, V) = \inf_{\rho \in \mathcal{C}_N} \mathcal{E}[\rho; B, V]. \quad (2.7)$$

Because of (2.4) the energy satisfies the scaling relation

$$E^{\text{MTF}}(N, B, V) = N^2 E^{\text{MTF}}(1, B/N, V/N). \quad (2.8)$$

In the limit $B \rightarrow 0$ the kinetic-energy density (2.3) converges to $j_0(\rho) = (\pi/2)\rho^2$ and (1.5) converges to the energy functional (1.3) of two-dimensional TF theory at $B=0$. It is easy to see that also $\lim_{B \rightarrow 0} E^{\text{MTF}}(N, B, V) = E^{\text{TF}}(N, V)$, where E^{TF} is defined in the same way as E^{MTF} with (1.3) replacing (1.5). We can thus consider the TF theory as a special case of MTF theory. In the opposite limit, $B \rightarrow \infty$, the kinetic-energy term vanishes altogether and one obtains a classical electrostatic model (1.6) that we shall study in Sec. III. Note also that since $j_B \leq j_0$ for all B it follows that $E^{\text{MTF}}(N, B, V) \leq E^{\text{TF}}(N, V)$ for all B . In particular $E^{\text{MTF}}(1, \beta, v)$ is uniformly bounded in the parameter $\beta = B/N$ for fixed $v = V/N$.

For fixed B and V , $E^{\text{MTF}}(N, B, V)$ is a convex, continuously differentiable function of N and, since $V \geq 0$, it is monotonically increasing. By the methods of Refs. 28, 29, and 21 (see also Ref. 30) it is straightforward to prove the existence and uniqueness of a minimizer for the variational problem (2.7).

2.1 THEOREM (minimizer). *There is a unique density $\rho_{N,B,V}^{\text{MTF}} \in \mathcal{C}_N$ such that*

$$E^{\text{MTF}}(N, B, V) = \mathcal{E}^{\text{MTF}}(\rho_{N,B,V}^{\text{MTF}}).$$

Note that the existence of a minimizing density with $\int \rho = N$ is guaranteed for all N because $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$. This condition on V also implies that $\rho_{N,B,V}$ vanishes outside a ball of finite radius, cf. Lemma A1 in the Appendix. The scaling relation for the minimizing density is

$$\rho_{N,B,V}^{\text{MTF}}(x) = N \rho_{1,B/N,V/N}^{\text{MTF}}(x). \quad (2.9)$$

Theorem 2.1 includes the TF theory as a special case. In the same way as in Prop. 4.14 in Ref. 21 one shows that $\rho_{N,B,V}^{\text{MTF}} \rightarrow \rho_{N,V}^{\text{TF}}$ weakly in L^1 as $B \rightarrow 0$.

The shape of the electronic density (computed by Kristinn Johnsen) in the case of a quadratic potential $V(x) = K|x|^2$ and $\gamma = 0$ is shown in Fig. 2 for different values of B . At the highest value of B (8 T), the density is everywhere below $d_0(B)$ and given by the minimizer (3.15) of the classical functional (1.6). At $B = 7$ T, all the electrons are still in the lowest Landau level, but that level is full around the middle of the dot where the density is anchored at $d_0(B)$. As the field gets weaker it becomes energetically favorable for electrons at the boundary of the dot, where the potential is high, to move into the next Landau level close to the minimum of the potential. A dome-shaped region then arises above the plateau at $\rho = d_0(B) = D_0(B)$, but eventually the density hits the next plateau at $\rho = D_1(B)$. This gradual filling of levels continues as the field strength goes down. At $B = 2$ T three Landau levels are full and electrons in the central dome are beginning to occupy the fourth level. Finally, at $B = 0$, we have the usual Thomas-Fermi model, which may be regarded as a limiting case with infinitely many Landau levels occupied.

In order to state the variational equation for the minimization problem it is convenient to define the derivative $j'_B = dj_B/d\rho$ of the kinetic-energy density everywhere, including points of discontinuity, as a *set valued* function (cf. Ref. 30), namely,

$$j'_B(\rho) = \begin{cases} \{\varepsilon_\nu(B)\} & \text{for } D_\nu(B) < \rho < D_{\nu+1}(B), \nu = 0, 1, \dots \\ [\varepsilon_\nu(B), \varepsilon_{\nu+1}(B)] & \text{for } \rho = D_{\nu+1}(B), \nu = 0, 1, \dots \end{cases} \quad (2.10)$$

With this notation the Thomas-Fermi equation for the functional (1.5) may be written as follows.

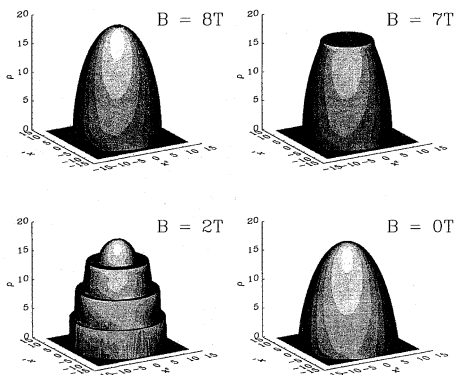


FIG. 2. Quantum dots at various magnetic field strengths. The potential is $V(x) = \frac{1}{2} m_* \omega^2 |x|^2$, with $m_* = 0.67m$, $\hbar\omega = 3.37$ meV, and $N = 50$. The coordinate axes are displayed in units of 10^{-8} m and the density ρ in the units 10^{-14} m $^{-2}$.

2.2 THEOREM (Thomas-Fermi equation). *There is a non-negative number $\mu = \mu(N, B, V)$ such that the minimizer $\rho = \rho_{N,B,V}^{\text{MTF}}$ satisfies*

$$\mu - V(x) - \rho^* |x|^{-1} \begin{cases} \in j'_B[\rho(x)] & \text{if } \rho(x) > 0 \\ \leq 0 & \text{if } \rho(x) = 0. \end{cases} \quad (2.11)$$

The quantity μ appearing in the TF equation is the physical chemical potential, i.e.,

$$\mu = \partial E(N, B, V) / \partial N. \quad (2.12)$$

Since E is convex as a function of N , μ is monotonically increasing with N for fixed B and V . It satisfies

$$\mu(N, B, V) = N \mu(1, B/N, V/N). \quad (2.13)$$

The derivation of the TF equation is analogous to that in Ref. 29. It is also true that if (ρ, μ) is *any* solution pair for (2.11), then ρ is the minimizer of $\mathcal{E}^{\text{MTF}}$ for some N and $\mu = \mu(N, B, V)$. The proof of this is a bit trickier than in the standard case,²⁹ because j_B is not continuously differentiable. It has been carried out by Lieb and Loss.³⁰

Finally we discuss the relationship between the MTF theory and the classical theory defined by the functional (1.6). We of course have that

$$\mathcal{E}^{\text{MTF}}[\rho; B, V] = \int j_B[\rho(x)] dx + \mathcal{E}[\rho; V].$$

From the definition of j_B one expects that the kinetic-energy term above can be neglected for large B and hence that $\lim_{B \rightarrow \infty} E^{\text{MTF}} = E^C$. The rigorous proof of this fact relies on a careful study of the classical problem. This analysis is far from trivial and is postponed to the next section.

There is another case where the MTF energy can be related to the classical energy. Namely, for a homogeneous exterior potential, i.e.,

$$V(\lambda x) = \lambda^s V(x)$$

for all $\lambda > 0$ with some $s > 0$. We consider the potentials $kV(x)$ with $k > 0$ and are interested in the dependence of the MTF energy and density on the coupling constant k .

Writing

$$\rho(x) = k^{2/(s+1)} \hat{\rho}(k^{1/(s+1)} x) \quad (2.14)$$

we have the scaling

$$\begin{aligned} \mathcal{E}^{\text{MTF}}[\rho; B, kV] &= k^{2/(s+1)} \int j_b(\hat{\rho}) + \mathcal{E}[\hat{\rho}; kV] \\ &= k^{1/(s+1)} \left[k^{1/(s+1)} \int j_b(\hat{\rho}) \right. \\ &\quad \left. + \mathcal{E}[\hat{\rho}; V] \right], \end{aligned} \quad (2.15)$$

where

$$b = B k^{-2/(s+1)}. \quad (2.16)$$

Changing k is thus equivalent to changing the kinetic energy by a multiplicative factor and rescaling the magnetic field, keeping the potential fixed. We shall show in the next section that for k small E^C is a good approximation to E^{MTF} .

III. THE CLASSICAL CONTINUOUS MODEL: A LIMIT OF MTF THEORY

For densities ρ small enough $[\rho(x) \leq d_1(B)]$ for all x the kinetic energy $j_B(\rho)$ vanishes. It is therefore natural to consider the resulting classical energy functional defined by (1.6), i.e.,

$$\mathcal{E}^C[\rho; V] = \int \rho(x) V(x) dx + \frac{1}{2} \iint \rho(x) |x - y|^{-1} \rho(y) dx dy. \quad (3.1)$$

The corresponding classical energy is

$$E^C(N, V) = \inf \left\{ \mathcal{E}^C[\rho; V] : \rho \geq 0, \int \rho = N \right\}. \quad (3.2)$$

$$E^C(N, V) = \inf \left\{ \mathcal{E}^C[\rho; V] : \text{support } \rho \subset \{x : |x| \leq R_v\}, \rho \geq 0, \int \rho = N \right\}. \quad (3.3)$$

Here R_v depends only on $v = V/N$. Later on we shall show that the minimizer is, indeed, a function, and hence that (3.2) does give us the large- B limit of MTF theory for suitable V .

3.1 PROPOSITION (existence and uniqueness of a minimizing measure). *Let V be continuous. Then there is a unique positive measure $\rho_{N,V}^C$ with $\int \rho_{N,V}^C = N$ such that $E^C(N, V) = \mathcal{E}^C[\rho_{N,V}^C; V]$.*

Proof. [Note: we write measures as $\rho(x)dx$, even if they are not absolutely continuous with respect to Lebesgue measure.] By (3.3) we can choose a sequence of posi-

In this section we analyze this functional and prove that it is, indeed, the large- B limit of MTF theory.

As before we assume that the confining potential V is positive and that $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$. Moreover, we shall here assume that V is continuous (in fact, we shall make an even more stringent regularity assumption in Theorem 3.2 below).

We begin by showing the existence of a minimizer for (3.2). For general continuous V (without further assumptions) we must take into account the possibility that the minimizing ρ may be a measure. In (3.2) we therefore minimize over all positive measures ρ with $\int \rho = N$. It follows from the finite radius lemma given in the Appendix that

tive measures, $\rho_1, \rho_2, \dots$, supported in $\{x : |x| \leq R_v\}$ with $\int \rho_n = N$, such that $\lim_{n \rightarrow \infty} \mathcal{E}^C[\rho_n; V] = E^C(N, V)$. The bounded measures are the dual of the continuous functions, and so, by the Banach-Alaoglu Theorem, we may assume (by possibly passing to a subsequence) that ρ_n converges weakly to a positive measure ρ still supported in $\{x : |x| \leq R_v\}$. In particular it follows that $\int \rho = N$ and $\int \rho_n V \rightarrow \int \rho V$. Moreover, the product measures $\rho_n \times \rho_n \rightarrow \rho \times \rho$ weakly. Hence

$$\begin{aligned} \iint \rho(x) \rho(y) |x - y|^{-1} dx dy &= \lim_{\delta \rightarrow 0} \iint \rho(x) \rho(y) (|x - y| + \delta)^{-1} dx dy \\ &= \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \iint \rho_n(x) \rho_n(y) (|x - y| + \delta)^{-1} dx dy \\ &\leq \liminf_{n \rightarrow \infty} \iint \rho_n(x) \rho_n(y) |x - y|^{-1} dx dy. \end{aligned} \quad (3.4)$$

The first equality follows by the Lebesgue's monotone convergence theorem. The last inequality is an immediate consequence of the pointwise bound $(|x - y| + \delta)^{-1} \leq |x - y|^{-1}$.

We conclude from (3.4) that

$$E^C(N, V) \leq \mathcal{E}^C[\rho; V] \leq \liminf_{n \rightarrow \infty} \mathcal{E}^C[\rho_n; V] = E^C(N, V), \quad (3.5)$$

and hence that ρ is a minimizer.

The uniqueness of ρ follows from strict convexity of $D(\rho, \rho)$. Q.E.D.

The next theorem gives conditions which are perfectly adequate for the physical applications under which the minimizer is a function and not just a measure. More-

over, that function has certain nice integrability properties.

3.2 THEOREM (the minimizer is a function). *Assume that the potential V is in the class $C_{loc}^{1,\alpha}$ for some $0 < \alpha \leq 1$ (i.e., V is once continuously differentiable and for each $R > 0$ its derivative satisfies*

$$|\nabla V(x) - \nabla V(y)| \leq c_R |x - y|^\alpha, \quad (3.6)$$

inside the ball of radius R centered at the origin for some constant $c_R > 0$). Then the minimizing measure $\rho_{N,V}^C$ of Proposition 3.1 is a function. It has the properties (with $\hat{p}$ being the Fourier transform of ρ)

$$\int |\hat{\rho}_{N,V}^C(p)|^2 |p|^r dp < C_1, \quad -1 \leq r < \alpha \quad (3.7)$$

$$\rho_{N,V}^C * |x|^{-1} \text{ is continuous,} \quad (3.8)$$

$$\int \rho_{N,V}^C(x)^q dx < C_2, \quad 1 \leq q < \frac{4}{2-\alpha}, \quad (3.9)$$

where C_1 and C_2 are constants (implicitly computed below) that depend only on the constants c_R, q, r, α and on N .

Proof. We write $\rho_{N,V}^C = \rho$. We know that $\int \rho = N$ and that ρ has compact support. From the former fact we conclude that $\hat{\rho}$ is well defined, continuous, and bounded.

Let g_a be the function with Fourier transform

$$\hat{g}_a(p) = \int g_a(x) e^{ipx} dx = \begin{cases} 1, & |p| \leq a \\ 0, & |p| \geq a. \end{cases} \quad (3.10)$$

Then g_a is continuous, $\int g_a = 1$, and $\int y g_a(y) dy = 0$.

Let ρ_a be the convolution $\rho * g_a$, so that $\int \rho_a = N$.

Since ρ is a minimizer, $\mathcal{E}^C[\rho; V] \leq \mathcal{E}^C[\rho_a; V]$. Explicitly this inequality is

$$\int (\rho_a - \rho) V + D(\rho_a, \rho_a) - D(\rho, \rho) \geq 0. \quad (3.11)$$

Since $\int V \rho_a = \int (V * g_a) \rho$ we can write the first term as

$$\int \int [V(x-y) - V(x)] \rho(x) g_a(y) dx dy.$$

By integrating ∇V along the line from $x-y$ to x , and using (3.6), we have

$$|V(x-y) - V(x) + y \cdot \nabla V(x)| \leq c |y|^{\alpha+1}.$$

(Note that by the finite radius Lemma A.1 all integrals are restricted to a finite ball.) Using the fact that $\int y g_a(y) dy = 0$ we can estimate the first term in (3.11) as follows:

$$\int (\rho_a - \rho) V \leq C \int |y|^{\alpha+1} |g_a(y)| dy = \text{const } a^{-\alpha-1}.$$

The last two terms in (3.11) are

$$\begin{aligned} \text{const } \int |\hat{\rho}(p)|^2 |p|^{-1} [|\hat{g}_a(p)|^2 - 1] dp \\ = \text{const } \int_{|p| \geq a} |\hat{\rho}(p)|^2 |p|^{-1} dp. \end{aligned}$$

(Recall that the Fourier transform of $|x|^{-1}$ is equal to $\text{const } |p|^{-1}$ in two dimensions.) The inequality (3.11) thus implies

$$\int_{|p| \geq a} |\hat{\rho}(p)|^2 |p|^{-1} dp \leq \text{const } a^{-\alpha-1}. \quad (3.12)$$

Using (3.12) and $\hat{\rho}(p) \leq \int \rho = N$ we can now prove (3.7) as follows:

$$\begin{aligned} \int |\hat{\rho}(p)|^2 |p|^r dp &= \int_{|p| \leq 1} |\hat{\rho}(p)|^2 |p|^r dp + \sum_{n=0}^{\infty} \int_{2^n \leq |p| \leq 2^{n+1}} |\hat{\rho}(p)|^2 |p|^r dp \\ &\leq N^2 \int_{|p| \leq 1} |p|^r dp + \text{const} \sum_{n=0}^{\infty} 2^{(r+1)(n+1)} \int_{2^n \leq |p| \leq 2^{n+1}} |\hat{\rho}(p)|^2 |p|^{-1} dp \\ &\leq (\text{const}) N^2 + \text{const} \sum_{n=0}^{\infty} 2^{(r+1)(n+1) - n(\alpha+1)} < \infty \end{aligned}$$

if $r < \alpha$.

To prove that $\rho * |x|^{-1}$ is continuous is now easy. We simply prove that its Fourier transform is integrable. The Fourier transform of $\rho * |x|^{-1}$ is $(\text{const}) \hat{\rho}(p) |p|^{-1}$ and

$$\int |\hat{\rho}(p)| |p|^{-1} dp \leq \int_{|p| \leq 1} |\hat{\rho}(p)| |p|^{-1} dp + \left[\int_{|p| \geq 1} |\hat{\rho}(p)|^2 |p|^r dp \right]^{1/2} \left[\int_{|p| \geq 1} |p|^{-r-2} dp \right]^{1/2} < \infty.$$

Finally, we prove (3.9). For $1 \leq q \leq 2$ there is no problem because we know that $\int \rho^2 = \int \hat{\rho}^2$. Hence ρ is a square integrable function and, since $\int \rho = N$, we conclude by Hölder's inequality that (3.9) holds for $1 \leq q \leq 2$. For $q > 2$ we will prove that

$$\int |\hat{\rho}|^t < \infty \quad \text{with} \quad \frac{4}{\alpha+2} < t = \frac{q}{q-1} < 2.$$

This will prove (3.9) by the Hausdorff-Young inequality, which states that $(\int |\hat{\rho}|^t)^{1/t} \geq (\int \rho^q)^{1/q}$ when $1 \leq t \leq 2$.

We write $|\hat{\rho}(p)|^t = \{|\hat{\rho}(p)|^t (1+|p|)^{ma}\} \{(1+|p|)^{-m}\}$ and then use Hölder's inequality with $a^{-1} + b^{-1} = 1$ to conclude that

$$\begin{aligned} \int |\hat{\rho}|^t &\leq \left[\int |\hat{\rho}(p)|^{ta} (1+|p|)^{ma} dp \right]^{1/a} \\ &\quad \times \left[\int (1+|p|)^{-mb} dp \right]^{1/b}. \end{aligned}$$

Thus $\int |\hat{\rho}|^t < \infty$ if we can satisfy $ta=2$, $ma < \alpha$, and $mb > 2$, in addition to $a^{-1} + b^{-1} = 1$. This requires $\alpha/a > m > 2/b$, or $1 \leq a < 1 + \frac{1}{2}\alpha$. Thus, we require $t=2/a > 4/(\alpha+2)$ which, since $q=t/(t-1)$, means $q < 4/(2-\alpha)$. Q.E.D.

Corresponding to the minimization (3.2) there is a variational equation satisfied by the minimizer ρ . In the general case in which ρ might be a measure, the variational equation exists but is slightly complicated to state.

In physically interesting cases V is certainly in $C^{1,\alpha}$, in which case Theorem 3.2 tells us that ρ is a function and that $\rho * |x|^{-1}$ is continuous. Hence the total potential

$$V^C = V + \rho * |x|^{-1} \quad (3.13)$$

is continuous. It is then easy to derive by standard arguments, as in Sec. I, that ρ is the unique non-negative solution to the variational equation

$$\begin{aligned} V(x) + \rho^* |x|^{-1} &= \mu \quad \text{if } \rho(x) \neq 0, \\ V(x) + \rho^* |x|^{-1} &\geq \mu \quad \text{if } \rho(x) = 0, \end{aligned} \quad (3.14)$$

for a unique $\mu > 0$. As usual the chemical potential μ is a monotone function of the particle number $N = \int \rho$. In the special case of a parabolic confining potential the solution to (3.14) can be given in closed form.

3.3 PROPOSITION (minimizer for the parabolic exterior potential). *If $V(x) = K|x|^2$ then the minimizer of $\mathcal{E}_{N,V}^C$ is*

$$\rho_{N,V}^C(x) = \begin{cases} \frac{3}{2\pi} N \lambda \sqrt{1 - \lambda|x|^2} & \text{if } |x| \leq \lambda^{-1} \\ 0 & \text{if } |x| > \lambda^{-1}, \end{cases} \quad (3.15)$$

where $\lambda = (8K/3\pi N)^{2/3}$. In fact, $\rho_{N,V}^C$ is the solution to (3.14) with $\mu = (3\pi/4)N\lambda^{1/2}$.

Proof. This solution (3.15) was certainly known before; see, e.g., Ref. 10. We give the proof here for the convenience of the reader. We only have to show that $\rho = \rho_{N,V}^C$ is the solution to (3.14). It is enough to consider the case $\lambda = 1$ and $N = 1$. Then $V(x) = (3\pi/8)|x|^2$ and $\mu = 3\pi/4$. We may compute $\rho^* |x|^{-1} = 3/2\pi \int \sqrt{1 - |x-y|^2} |y|^{-1} dy$ by writing y in polar coordinates $(|y|, \theta)$ and performing the $|y|$ integration first:

$$\begin{aligned} \rho^* |x|^{-1} &= \frac{3}{2\pi} \iint [(1 - |y| - |x|\cos\theta)^2 \\ &\quad - |x|^2 \sin^2\theta]^{1/2} d\theta d|y| \\ &= \frac{3}{2\pi} \int (1 - |x|^2 \sin^2\theta)^{1/2} \\ &\quad \times \int \left[1 - \frac{(|y| - |x|\cos\theta)^2}{1 - |x|^2 \sin^2\theta} \right]^{1/2} d|y| d\theta, \end{aligned}$$

the integrations are over the intervals in θ and $|y|$ for which the integrands are real. Introducing the variable $t = (|y| - |x|\cos\theta)(1 - |x|^2 \sin^2\theta)^{-1/2}$ we obtain

$$\begin{aligned} \rho^* |x|^{-1} &= \frac{3}{2\pi} \int_{-\theta_m(x)}^{\theta_m(x)} (1 - |x|^2 \sin^2\theta) d\theta \int_{-1}^1 (1 - t^2)^{1/2} dt \\ &= \frac{3}{4} \int_{-\theta_m(x)}^{\theta_m(x)} (1 - |x|^2 \sin^2\theta) d\theta, \end{aligned} \quad (3.16)$$

where

$$\theta_m(x) = \begin{cases} \pi/2 & \text{if } |x| \leq 1 \\ \sin^{-1} \frac{1}{|x|} & \text{if } |x| > 1. \end{cases}$$

Thus

$$\rho^* |x|^{-1} = \frac{3\pi}{4} - \frac{3\pi}{8} |x|^2 \quad \text{if } |x| \leq 1$$

and

$$\rho^* |x|^{-1} \geq \frac{3\pi}{4} - \frac{3\pi}{8} |x|^2 \quad \text{if } |x| > 1.$$

[The last inequality comes from the fact that the integral

from θ_m to $\pi/2$ in (3.16) is negative.] Q.E.D.

The energy function $E^C(N, V)$ has the simple scaling:

$$E^C(N, V) = N^2 E^C \left[1, \frac{V}{N} \right]. \quad (3.17)$$

The minimizing density $\rho_{N,V}^C$ for (3.2) scales as

$$\rho_{N,V}^C(x) = N \rho_{1,v}^C(x)$$

where $v = V/N$.

We shall now make precise in what sense the classical energy E^C is a limit of the MTF energy. In fact, in two different limits (the large- B limit and the low coupling limit) the MTF energy will converge to the classical energy. We first treat the large- B limit.

3.4 THEOREM (large- B limit of MTF). *If the exterior potential V is in the class $C_{\text{loc}}^{1,\alpha}$ we have, as $B \rightarrow \infty$,*

$$E^{\text{MTF}}(N, B, V) \rightarrow E^C(N, V) \quad (3.18)$$

and

$$\rho_{N,B,V}^{\text{MTF}}(x) \rightarrow \rho_{N,V}^C, \quad (3.19)$$

in the weak L^1 sense.

Proof. If we use $\rho_{N,B,V}^{\text{MTF}}$ as a trial density in $\mathcal{E}^C$ and recall that $j_B \geq 0$ we immediately obtain $E^C(N, V) \leq E^{\text{MTF}}(N, B, V)$.

For the bound in the opposite direction we use $\rho_{N,V}^C$ as a trial density for $\mathcal{E}^{\text{MTF}}$. In order to do this it is, however, important that we know (from Theorem 3.2) that $\rho_{N,V}^C$ is a function. Hence $j_B(\rho_{N,V}^C)$ is well defined. Moreover, from the definition of j_B , $j_B(\rho_{N,V}^C) \rightarrow 0$ almost everywhere as $B \rightarrow \infty$ and $j_B(\rho_{N,V}^C) \leq j_0(\rho_{N,V}^C)$. Since $j_0(\rho) = (\pi/2)(\rho)^2$ we know from (3.9) that $j_0(\rho_{N,V}^C)$ is integrable. The limit in (3.18) is therefore an immediate consequence of Lebesgue's dominated convergence theorem.

The convergence of the densities in (3.19) follows in a standard way by replacing V by $V + \epsilon f$ with f a bounded (measurable) function and differentiating with respect to ϵ , see e.g., Ref. 29. Q.E.D.

We point out that if $\rho_{N,V}^C$ is a bounded function, as it is, e.g., for $V = K|x|^2$, then

$$E^{\text{MTF}}(N, B, V) = E^C(N, V), \quad (3.20)$$

for B large enough, because in that case $j_B(\rho_{N,V}^C)$ vanishes for B large.

Finally, we now discuss the weak-coupling limit in the case of homogeneous exterior potentials. Suppose V is a homogeneous function of x , $V(\lambda x) = \lambda^s V(x)$, $s > 0$. If we consider the exterior potentials $kV(x)$ with $k > 0$ the classical energy and density obey the scalings

$$E^C(N, kV) = k^{1/(s+1)} E^C(N, V) \quad (3.21)$$

and

$$\rho_{N,kV}^C(x) = k^{2/(s+1)} \rho_{N,V}^C(k^{1/(s+1)} x). \quad (3.22)$$

If k is small we see from (3.22) that the minimizing density for the MTF functional will spread out and its kinetic energy will be negligible compared with the classi-

cal terms. We prove this rigorously now.

3.5 THEOREM (weak-coupling limit of MTF with homogeneous potentials). Let V be $C_{\text{loc}}^{1,\alpha}$ and homogeneous of degree s . If $k \rightarrow 0$ then

$$\frac{E^{\text{MTF}}(N, B, kV)}{E^C(N, kV)} \rightarrow 1 \quad (3.23)$$

and

$$k^{-2/(s+1)} \rho_{N,B,kV}^{\text{MTF}}(k^{-1/(s+1)}x) \rightarrow \rho_{N,V}^C(x), \quad (3.24)$$

in weak L^1 sense. Both limits are uniform in B .

Proof. As above we may use $\rho_{N,B,V}^{\text{MTF}}$ as a trial density in $\mathcal{E}^C$ to conclude that $E^C(N, V) \leq E^{\text{MTF}}(N, B, V)$.

To prove the bound in the opposite direction we again use $\rho_{N,V}^C$ as a trial density for $\mathcal{E}^{\text{MTF}}$. We then obtain from (2.15) and the scaling (3.22) that

$$E^{\text{MTF}}(N, B, kV) \leq k^{2/(s+1)} \int j_0(\rho_{N,V}^C) + E^C(N, kV), \quad (3.25)$$

where we used that $j_b \leq j_0$. If we compare this with the scaling in (3.21) we see that $E^{\text{MTF}}/E^C \rightarrow 1$ as $k \rightarrow 0$ since $j_0(\rho_{N,V}^C) = (\pi/2)(\rho_{N,V}^C)^2$ is integrable.

The convergence of the densities follows again by replacing V by $V + \epsilon f$ and differentiating with respect to ϵ . Q.E.D.

In the same way as for the large- B limit (3.23) becomes an identity for small k if $\rho_{N,V}^C$ is a bounded function.

We may of course also introduce the scaling $V = Nv$ when v is homogeneous of degree s . Then $kV = Kv$, where $K = kN$, and the limit in (3.23) is uniform in N . The limit in (3.24) is uniform if we formulate it as

$$N^{-1} k^{-2/(s+1)} \rho_{N,B,kV}^{\text{MTF}}(k^{-1/(s+1)}x) \rightarrow \rho_{1,v}^C(x). \quad (3.26)$$

We remark that if a potential W is asymptotically homogeneous in the sense that there is a homogeneous potential V with $\lim_{|x| \rightarrow \infty} W(x)/V(x) = 1$, then

$$\lim_{k \rightarrow 0} E(N, B, kW)/k^{1/(1+s)} = E^C(N, V) \quad (3.27)$$

uniformly in B , where s is the degree of homogeneity of V .

IV. THE CLASSICAL POINT CHARGE MODEL: A LIMIT OF QUANTUM MECHANICS

Another model that sheds some light on the physics of our problem—and that will also be important for bounding the difference between the TF theory and the original quantum theory in Sec. V—is the classical particle model. In this model the kinetic energy is simply omitted altogether, but the pointlike nature of the electrons is retained.

4.1 DEFINITION (classical particle energy). With $V(x)$ being the confining potential the classical particle energy for N points in $\mathbb{R}^2$ is defined by

$$\mathcal{E}^P(x_1, \dots, x_N; V) = \sum_{i=1}^N V(x_i) + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}. \quad (4.1)$$

The minimum classical particle energy for N point particles in $\mathbb{R}^2$ is

$$E^P(N, V) = \inf \{ \mathcal{E}^P(x_1, \dots, x_N; V) : x_i \in \mathbb{R}^2 \}. \quad (4.2)$$

We shall estimate the particle energy $E^P(N, V)$ in terms of the classical continuum energy $E^C(N, V)$. We first show that $E^C(N, V)$ gives an exact upper bound on $E^P(N, V)$.

4.2 LEMMA (upper bound for E^P). For all N we have

$$E^P(N, V) \leq E^C(N, V) - N^{3/2}/(8R_v), \quad (4.3)$$

where R_v is the maximal radius given in Lemma A1.

Proof. First, let us give a very simple argument that yields an error term proportional to N instead of $N^{3/2}$. The energy $E^P(N, V)$ is bounded above by

$$\int \mathcal{E}^P(x_1, \dots, x_N; V) \Phi(x_1, \dots, x_N) dx_1 \cdots dx_N,$$

for any non-negative function Φ with $\int \Phi = 1$. We take $\Phi(x_1, \dots, x_N) = \prod_{i=1}^N \rho_{[1]}(x_i)$, where for simplicity we have introduced the notation $\rho_{[1]}$ for the minimizer $\rho_{1,V/N}^C$ for $\mathcal{E}^C[\rho; V/N]$ with $\int \rho_{[1]}^C = 1$. Note that $\rho_{[1]}$ depends only on $v = V/N$. We obtain

$$\begin{aligned} E^P(N, V) &\leq \int \mathcal{E}^P(x_1, \dots, x_N; V) \prod_{i=1}^N \rho_{[1]}(x_i) dx_1 \cdots dx_N \\ &= N \int V(x) \rho_{[1]}(x) dx \\ &\quad + \frac{N(N-1)}{2} \int \int \rho_{[1]}(x) |x-y|^{-1} \\ &\quad \times \rho_{[1]}(y) dx dy. \end{aligned}$$

Recalling that the minimizer of $\mathcal{E}^C$ is $\rho_{N,V}^C(x) = N\rho_{[1]}(x)$, we get an error term $-aN$, with $a = \frac{1}{2} \int \int \rho_{[1]}(x) \rho_{[1]}(y) |x-y|^{-1} dx dy$.

Now we turn to a proof of (4.3) which, obviously, has to be more complicated than the previous discussion. By Lemma A1 there is a fixed square Q centered at the origin, whose width W equals $2R_v$, such that the minimizer $\rho = \rho_{N,V}^C$ for $\mathcal{E}^C$ is supported in Q . For simplicity we suppose that $\sqrt{N}$ is an integer; if this is not so the following proof can be modified in an obvious way.

First, cut Q into $\sqrt{N}$ vertical, disjoint strips, $S_1, S_2, \dots, S_{\sqrt{N}}$ such that $\int_{S_j} \rho = \sqrt{N}$ for all j . Let t_j denote the width of S_j , so that $\sum_{j=1}^{\sqrt{N}} t_j = W$. Next, make $\sqrt{N} - 1$ horizontal cuts in each S_j so that the resulting rectangles R_{jk} for $k=1, \dots, \sqrt{N}$ satisfy $\int_{R_{jk}} \rho = 1$. Denote the height of these rectangles by h_{jk} , so that $\sum_{k=1}^{\sqrt{N}} h_{jk} = W$ for each j . Having done this we note, by convexity, that for each j

$$\begin{aligned} N^{-1/2} \sum_{k=1}^{\sqrt{N}} (t_j + h_{jk})^{-1} &\geq \left[N^{-1/2} \sum_{k=1}^{\sqrt{N}} (t_j + h_{jk}) \right]^{-1} \\ &= [t_j + N^{-1/2} W]^{-1}. \end{aligned}$$

Again, using the same convexity argument for the j summation, we have that

$$\sum_{j=1}^{\sqrt{N}} \sum_{k=1}^{\sqrt{N}} (t_j + h_{jk})^{-1} \geq \frac{N^{3/2}}{2W}. \quad (4.4)$$

Let ρ_{jk} be the minimizing density ρ restricted to the rectangle R_{jk} , i.e., $\rho_{jk}(x)=1$ if $x \in R_{jk}$ and $=0$ otherwise. Thus, $\int \rho_{jk} = 1$. We denote these N functions by ρ^i , $i=1, \dots, N$. Define $\Phi(x_1, \dots, x_N) := \prod_{i=1}^N \rho^i(x_i)$ and, as in the previous proof, a simple computation yields

$$E^P(N, V) \leq \int \mathcal{E}^P \Phi = \mathcal{E}^C[\rho; V] - \sum_{i=1}^N D(\rho^i, \rho^i),$$

with $D(f, g) = \frac{1}{2} \int \int f(x)g(y)|x-y|^{-1} dx dy$.

To complete our proof we note that as long as x and y are in R_{jk} we have that $|x-y|^{-1} \geq (t_j + h_{jk})^{-1}$. Thus, $\sum_{j,k} D(\rho_{jk}, \rho_{jk}) \geq N^{3/2}/4W$ by (4.4) and the fact that $\int \rho_{jk} = 1$. Q.E.D.

4.3 LEMMA (lower bound for E^P). Assume that V is a potential in $C_{loc}^{1,\alpha}$. Then for all N we have

$$E^P(N, V) \geq E^C(N, V) - bN^{3/2}. \quad (4.5)$$

with

$$b = \frac{4}{3} \sqrt{\frac{1}{3}} \int [\rho_{1,v}^C(x)]^{3/2} dx + \left[\frac{2\pi}{2-p} \right]^{1/p} (2R_v)^{-1+(2/p)} \left[\int (\rho_{1,v}^C)^q \right]^{1/q} \quad (4.6)$$

and where q is any number satisfying $2 < q < 4/(2-\alpha)$, R_v is the maximal radius given in Lemma A1 and $p = q/(q-1) < 2$. As we explained in Theorem 3.2, our hypothesis that $V \in C_{loc}^{1,\alpha}$ implies that for $q < 4/(2-\alpha)$, $[\int (\rho_{1,v}^C)^q]^{1/q}$ is less than some constant that depends only on q , α , and $v = V/N$. In particular, b depends only on v .

We see from (4.3) and (4.5) that when $V \in C_{loc}^{1,\alpha}$ the power $\frac{3}{2}$ in the error term is optimal.

In order to prove (4.5) we need the following electro-

statics lemma of Lieb and Yau.²⁶ The original version was for $\mathbb{R}^3$; we state it here for $\mathbb{R}^2$ solely for the convenience of our present application.

4.4 LEMMA (the interaction of points and densities). Given points $x_1, \dots, x_N$ in $\mathbb{R}^2$, we define Voronoi cells $\Gamma_1, \dots, \Gamma_N \subset \mathbb{R}^2$ by

$$\Gamma_j = \{y \in \mathbb{R}^2 : |y-x_j| \leq |y-x_k| \text{ for all } k \neq j\}.$$

These Γ_j have disjoint interiors and their union covers $\mathbb{R}^2$. We also define R_j to be the distance from x_j to the boundary of Γ_j , i.e., R_j is half the distance of x_j to its nearest neighbor. Let ρ be any (not necessarily positive) function on $\mathbb{R}^2$. (In general, ρ can be replaced by a measure, but it is not necessary for us to do so.) Then [with $D(f, g) = \frac{1}{2} \int \int f(x)g(y)|x-y|^{-1} dx dy$]

$$\begin{aligned} \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} &\geq -D(\rho, \rho) \\ &+ \sum_{j=1}^N \int_{\mathbb{R}^2} \rho(y) |y - x_j|^{-1} dy \\ &+ \frac{1}{8} \sum_{j=1}^N R_j^{-1} \\ &- \sum_{j=1}^N \int_{\Gamma_j} \rho(y) |y - x_j|^{-1} dy. \end{aligned} \quad (4.7)$$

Proof of (4.5). We choose ρ in (4.7) to be the minimizer $\rho_{N,V}^C$ for the functional $\mathcal{E}^C$ and we choose the x_i 's to be any (not necessarily minimizing) configuration for $\mathcal{E}^P$. It is important for us that a minimizer exists for $\mathcal{E}^C$, for then ρ satisfies the Euler-Lagrange equation (3.14). Since $\int \rho = N$, we conclude from (3.14) that

$$\sum_{j=1}^N \left[\int \rho(y) |y - x_j|^{-1} dy + V(x_j) \right] = \sum_{j=1}^N V^C(x_j) \geq N\mu = \int V^C \rho = 2D(\rho, \rho) + \int V \rho.$$

Thus, if we add $\sum_j V(x_j)$ to both sides of (4.7), we have that

$$E^P(N, V) \geq E^C(N, V) + \sum_{j=1}^N \frac{1}{8} R_j^{-1} - \int_{\Gamma_j} \rho(y) |y - x_j|^{-1} dy. \quad (4.8)$$

Our goal will be to control the rightmost term in (4.8) by the R_j^{-1} term.

We split each region Γ_j into two disjoint subregions, $\Gamma_j = A_j \cup B_j$, where

$$A_j := \{x : |x - x_j| < R_j\}, \quad B_j := \{x \in \Gamma_j : |x - x_j| \geq R_j\}.$$

Then, by Hölder's inequality

$$\begin{aligned} \sum_{j=1}^N \int_{B_j} |y - x_j|^{-1} \rho(y) dy &\leq \left[\sum_j \int_{\Gamma_j} \rho^{3/2} \right]^{2/3} \left[\sum_j \int_{|y-x_j| \geq R_j} |y - x_j|^{-3} dy \right]^{1/3} \\ &= \left[\int \rho^{3/2} \right]^{2/3} \left[2\pi \sum_j R_j^{-1} \right]^{1/3}. \end{aligned} \quad (4.9)$$

If we define $X := \sum_j R_j^{-1}$ we can rewrite (4.9) plus the $X/8$ term in (4.8) as $\frac{1}{4}X - (\int \rho^{3/2})^{2/3} X^{1/3}$. The minimum of this quantity, over all values X , is $\frac{4}{3}\sqrt{2/3} \int \rho^{3/2}$, and thus we have accounted for the first error term in (4.6).

To estimate the term $I := \sum_{j=1}^N \int_{A_j} \rho(y) |y - x_j|^{-1} dy$ some control is needed over the possible singularities of ρ . Let $p = q/(q-1)$ be the dual of q . Then

$$\begin{aligned} I &\leq \left[\int \rho^q \right]^{1/q} \left[\sum_j \int_{A_j} |y - x_j|^{-p} dy \right]^{1/p} \\ &= \left[\int \rho^q \right]^{1/q} \left[\frac{2\pi}{2-p} \sum_j R_j^{2-p} \right]^{1/p}. \end{aligned} \quad (4.10)$$

We note that, since $1 \leq p < 2$,

$$\left[\sum_{j=1}^N R_j^{2-p} \right]^{1/p} \leq \left[\sum_{j=1}^N R_j^2 \right]^{1/p-1/2} N^{1/2}$$

by Hölder's inequality. Now πR_j^2 is the area of the disc A_j and thus $\pi \sum_j R_j^2$ is the total area of all these disjoint discs. How large can this area be? To answer this we recall Lemma A1 in the Appendix, which states that for the purpose of finding a set of points that minimizes the classical particle energy $\mathcal{E}^p$ we can restrict attention to a disc of radius R_v , centered at the origin. We may therefore assume that our x_j 's satisfy $|x_j| \leq R_v$. This we can do whether or not an energy minimizing configuration exists. Having done so and assuming that $N \geq 2$ we have that $R_j < R_v$ for all j , and hence all our discs are contained in a disc of radius $2R_v$ centered at the origin. Thus $\sum_j R_j^2 \leq (2R_v)^2$, and our second error term, (4.10), is bounded above by

$$\left[\int \rho^q \right]^{1/q} \left[\frac{2\pi}{2-p} \right]^{1/p} (2R_v)^{-1+2/p} N^{1/2}.$$

This yields (4.6). Q.E.D.

V. MTF THEORY IS THE HIGH-DENSITY LIMIT OF QUANTUM MECHANICS

In this section we prove that the quantum energy and the quantum density are given by the corresponding MTF quantities to leading order for large N . These are the statements of Theorems 1.1 and 1.2. We shall not prove Theorem 1.2 since it follows from Theorem 1.1 in a standard way by replacing V by $V + \epsilon f$ with f a function in $C_{loc}^{1,\alpha}$ and differentiating w.r.t. ϵ , see, e.g., Ref. 29.

We shall prove Theorem 1.1 by giving sharp upper and lower bounds to the quantum ground-state energy. The upper bound is obtained by a variational calculation using the magnetic coherent states introduced in Refs. 25 and 21. The lower bound is more difficult. Besides the results of the previous sections several ingredients are needed. The first ingredient is a kinetic energy inequality of the Lieb-Thirring type.^{31,32,21,33} Such an inequality estimates the kinetic energy of a many-body wave function from below in terms of a functional of the density. The proof of this inequality in the two-dimensional case considered here is harder than the three-dimensional case

treated in Ref. 21 and involves some new mathematical ideas. The second ingredient is a lower bound on the exchange-correlation energy. The proof of this inequality is similar to that given in Refs. 34 and 35 for the three-dimensional case.

Once the kinetic energy and exchange-correlation inequalities have been established the proof of the lower bound is completed by a coherent states analysis.

We start by discussing the magnetic coherent states used in the proofs of both the upper and lower bounds. They are constructed from the kernels

$$\begin{aligned} \Pi_{\alpha\sigma}(x\sigma', y\sigma'') &= \frac{B}{2\pi} \exp[i(x \times y) \cdot B - |x - y|^2 B/4] \\ &\quad \times L_\alpha(|x - y|^2 B/2) \delta_{\alpha\sigma'} \delta_{\sigma\sigma''} \end{aligned} \quad (5.1)$$

of the projection operators onto the Landau levels $\alpha = 0, 1, 2, \dots$ with z -component of spin $\sigma = \pm \frac{1}{2}$. Here L_α are Laguerre polynomials normalized by $L_\alpha(0) = 1$. In fact, all that matters are the projectors Π_ν on the states with energy $\epsilon_\nu(B)$; these are given by a sum of at most two of the projections $\Pi_{\alpha\sigma}$. More precisely,

$$\Pi_\nu = \sum_{\alpha, \sigma} \Pi_{\alpha\sigma} \quad \text{with } \alpha + \frac{1}{2} + \gamma\sigma = \epsilon_\nu(B)/B. \quad (5.2)$$

We shall not need the explicit form (5.1). The three important properties of Π_ν that we use are the following:

$$\sum_\nu \Pi_\nu(x\sigma', y\sigma'') = \delta(x - y) \delta_{\sigma'\sigma''}, \quad (5.3)$$

$$\sum_{\sigma'} \Pi_\nu(x\sigma', x\sigma') = d_\nu(B), \quad (5.4)$$

$$H_{\text{kin}} \Pi_\nu = \epsilon_\nu(B) \Pi_\nu, \quad (5.5)$$

where H_{kin} is given by (2.1).

Let g be a real continuous function on $\mathbb{R}^2$, with $g(x) = 0$ for $|x| > 1$, $\int g^2 = 1$, and $\int (\nabla g)^2 < \infty$. [The optimal choice that minimizes $\int (\nabla g)^2$ is the Bessel function J_0 , suitably scaled and normalized.] Define $g_r(x) = r^{-1} g(x/r)$, with $0 < r < 1$ to be specified later. For each $u \in \mathbb{R}^2$, $\nu = 0, 1, 2$, we define the operator $\Pi_{\nu u}$ — the coherent “operator” — with kernel

$$\Pi_{\nu u}(x\sigma', y\sigma'') = g_r(x - u) \Pi_\nu(x\sigma', y\sigma'') g_r(y - u). \quad (5.6)$$

It easily follows from (5.3) and (5.4) and the properties of g that these kernels satisfy the coherent operator identities³⁶

$$\sum_\nu \int \Pi_{\nu u}(x\sigma', y\sigma'') du = \delta(x - y) \delta_{\sigma'\sigma''}, \quad (5.7)$$

$$\text{Tr} \Pi_{\nu u} = \sum_\sigma \int \Pi_{\nu u}(x\sigma', x\sigma') dx = d_\nu(B). \quad (5.8)$$

Moreover, a simple computation gives, using (5.5),

$$\text{Tr}[H_{\text{kin}} \Pi_{\nu u}] = d_\nu(B) \left[\epsilon_\nu(B) + \int (\nabla g_r)^2 \right], \quad (5.9)$$

$$\text{Tr}[V \Pi_{\nu u}] = d_\nu(B) V * g_r^2(u), \quad (5.10)$$

where V is a (continuous) potential and $*$ denotes convolution. Likewise, for all f with $\langle f|f \rangle = 1$

$$\langle f | H_{\text{kin}} | f \rangle = \sum_v \int \varepsilon_v(B) \langle f | \Pi_{v,u} | f \rangle du - \int (\nabla g_r)^2, \quad (5.11)$$

$$\langle f | V * g_r^2 | f \rangle = \sum_v \int V(u) \langle f | \Pi_{v,u} | f \rangle du. \quad (5.12)$$

Equations (5.9) and (5.10) will be used in proving the upper bound, while (5.11) and (5.12) are needed for the lower bound.

A. The upper bound

We use the variational principle of Ref. 37. According to this principle

$$E^Q(N, B, V) \leq \text{Tr}[(H_{\text{kin}} + V)K] + \frac{1}{2} \sum_{\sigma, \sigma'} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{K(x\sigma, x\sigma') K(y\sigma', y\sigma')}{|x - y|} \times dx dy \quad (5.13)$$

$$f_v(x) = \begin{cases} 1, & v \leq v_{\max}(x) \\ [\rho^{\text{MTF}}(x) - \sum_{v \leq v_{\max}(x)} d_v(B)] / d_{v_{\max}(x)+1}(B), & v = v_{\max}(x) + 1 \\ 0, & v > v_{\max}(x) + 1 \end{cases} \quad (5.17)$$

and define

$$K(x\sigma, y\sigma') = \sum_v \int f_v(u) \Pi_{v,u}(x\sigma, y\sigma') du, \quad (5.18)$$

with $\Pi_{v,u}$ as in (5.6). It follows from (5.7) that K satisfies (5.14) and from (5.8), (5.16), and (5.17) that $\text{Tr}[K] = \int \rho^{\text{MTF}}(u) du = N$.

Note that (5.4) and (5.6) imply

$$\sum_{\sigma} K(x\sigma, x\sigma) = \rho^{\text{MTF}} * g_r^2(x). \quad (5.19)$$

Hence, the last term in (5.13) is $D(\rho^{\text{MTF}} * g_r^2, \rho^{\text{MTF}} * g_r^2)$, where the functional D was defined in (1.4). By convexity of D we find that

$$D(\rho^{\text{MTF}} * g_r^2, \rho^{\text{MTF}} * g_r^2) \leq D(\rho^{\text{MTF}}, \rho^{\text{MTF}}).$$

From (5.9), (5.10), and (5.13) we obtain

$$E^Q(N, B, V) \leq \mathcal{E}^{\text{MTF}}(\rho^{\text{MTF}}) + N \int (\nabla g_r)^2 dx + \int [V * g_r^2(*) - V(x)] \rho^{\text{MTF}}(x) dx \leq E^{\text{MTF}}(N, V, B) + Nr^{-2} \int [\nabla g(x)]^2 dx + N^2 \sup_{|x| < R} [v * g_r^2(x) - v(x)], \quad (5.20)$$

where $R = R_v$ is the finite radius and we have written $V = Nv$. Since v is in $C^{1,\alpha}$

$$\sup_{|x| < R} |v * g_r^2(x) - v(x)| \leq (\text{const})r.$$

for all operators K , with kernel $K(x\sigma, y\sigma')$, satisfying

$$0 \leq \langle f | K | f \rangle \leq \langle f | f \rangle \quad (5.14)$$

for all f , and

$$\text{Tr}[K] = \sum_{\sigma} \int_{\mathbb{R}^2} K(x\sigma, x\sigma) dx = N. \quad (5.15)$$

We shall choose K as follows. Let ρ^{MTF} be the MTF density, i.e., the minimizer of the functional (1.5) with $\int \rho^{\text{MTF}} = N$. Denote by $v_{\max}(x)$ the highest filled level. Then

$$0 \leq \rho^{\text{MTF}}(x) - \sum_{v \leq v_{\max}(x)} d_v(B) < d_{v_{\max}(x)+1}(B). \quad (5.16)$$

We introduce the filling factors

We can choose $r = r_N$ such that $r_N \rightarrow 0$ and $r_N^{-2}/N \rightarrow 0$ as $N \rightarrow \infty$. This means that r_N should go to zero but still be large compared with the average spacing $N^{-1/2}$ between electrons. The optimal choice is of the order $r = (\text{const})N^{-1/3}$. Thus the error

$$[E^Q(N, B, V) - E^{\text{MTF}}(N, B, V)]N^{-2}$$

is bounded above by a function $\varepsilon_N^+(v) = c^+(v)N^{-1/3}$ (independent of B). This finishes the proof of the upper bound.

5.1 THEOREM (Lieb-Thirring inequality in two dimensions). Let $H_A = \frac{1}{2}(\nabla - \mathbf{A})^2 + \mathbf{S} \cdot \mathbf{B}$. [This is the operator H_{kin} from (2.1) with $\gamma = 1$.] Let W be a locally integrable function and denote by $e_1(W), e_2(W), \dots$ the negative eigenvalues (if any) of the operator $H = H_A - W$ defined on $L^2(\mathbb{R}^2; \mathbb{C}^2)$, the space of wave functions of a single spin- $\frac{1}{2}$ particle. Define $|W|_+(x) = \frac{1}{2}[|W(x)| + W(x)]$. For all $0 < \lambda < 1$ we then have the estimate

$$\sum_j |e_j(W)| \leq \lambda^{-1} \frac{B}{2\pi} \int_{\mathbb{R}^2} |W|_+(x) dx + \frac{3}{4}(1 - \lambda)^{-2} \int_{\mathbb{R}^2} |W|_+^2(x) dx.$$

Proof. For any self-adjoint operator A we denote by $N_a(A)$ the number of eigenvalues of A greater than or equal to a .

Since replacing W by its positive part $|W|_+$ will only enhance the sum of the negative eigenvalues we shall henceforth assume that W is positive, i.e., $W = |W|_+$.

We consider the Birman-Schwinger kernel

$$K_E = W^{1/2}(H_A + E)^{-1}W^{1/2}.$$

According to the Birman-Schwinger principle (see, e.g., Ref. 38, p. 89) the number $N_E(-H)$ of eigenvalues of H below $-E$ is equal to the number $N_1(K_E)$ of eigenvalues of K_E greater than or equal to 1. We find

$$\sum_j |e_j(W)| = \int_0^\infty N_E(-H)dE = \int_0^\infty N_1(K_E)dE.$$

In order to estimate $N_1(K_E)$ we decompose the Birman-Schwinger kernel into a part K_E^0 coming from the lowest Landau level and a part $K_E^>$ coming from the higher levels. If Π_0 is the projection onto the lowest Landau band these two parts are defined by

$$K_E^0 = W^{1/2}\Pi_0(H_A + E)^{-1}\Pi_0W^{1/2} = E^{-1}W^{1/2}\Pi_0W^{1/2} \quad (5.21)$$

and

$$K_E^> = W^{1/2}(I - \Pi_0)(H_A + E)^{-1}(I - \Pi_0)W^{1/2}.$$

Since Π_0 commutes with H_A we have $K_E = K_E^0 + K_E^>$.

Now we use Fan's theorem,³⁹ which states that if $\mu_1(X) \geq \mu_2(X) \geq \dots$ denote the eigenvectors of a self-adjoint compact operator X then $\mu_{n+m+1}(X+Y) \leq \mu_{n+1}(X) + \mu_{m+1}(Y)$ for $n, m \geq 0$. From this we have $N_1(X+Y) \leq N_\lambda(X) + N_{1-\lambda}(Y)$ which, in our case, reads as follows:

$$N_1(K_E) \leq N_\lambda(K_E^0) + N_{1-\lambda}(K_E^>) \quad \text{for } 0 \leq \lambda \leq 1.$$

This inequality permits us to consider the two parts of K_E separately. We first consider the contribution from the lowest level: $N_\lambda(K_E^0) = N_{\lambda E}(W^{1/2}\Pi_0W^{1/2})$. We get

$$\begin{aligned} \int_0^\infty N_\lambda(K_E^0)dE &= \int_0^\infty N_{\lambda E}(W^{1/2}\Pi_0W^{1/2})dE \\ &= \lambda^{-1} \int_0^\infty N_E(W^{1/2}\Pi_0W^{1/2})dE \\ &= \lambda^{-1} \text{Tr}(W^{1/2}\Pi_0W^{1/2}) \\ &= \lambda^{-1} \int \Pi_0(x, x)W(x)dx \\ &= \lambda^{-1} \frac{B}{2\pi} \int W(x)dx. \end{aligned}$$

The second part is straightforward. We first notice that $H_A(I - \Pi_0) \geq B(I - \Pi_0)$. Hence $H_A(I - \Pi_0) \geq \frac{2}{3}(H_A + \frac{1}{2}B)(I - \Pi_0) \geq \frac{1}{3}(i\nabla - A)^2(I - \Pi_0)$. [Note that $(i\nabla - A)^2$ commutes with Π_0 .] Since the operator inequality $0 < X \leq Y$ implies $X^{-1} \leq Y^{-1}$ we have that

$$\begin{aligned} K_E^> &\leq W^{1/2}(I - \Pi_0)[\tfrac{1}{3}(i\nabla - A)^2 + E]^{-1}(I - \Pi_0)W^{1/2} \\ &\leq W^{1/2}[\tfrac{1}{3}(i\nabla - A)^2 + E]^{-1}W^{1/2}. \end{aligned}$$

We conclude that $N_{1-\lambda}(K_E^>) \leq N_1(\tilde{K}_E)$, where

$$\begin{aligned} \tilde{K}_E &= [(1-\lambda)^{-1}W]^{1/2}[\tfrac{1}{3}(i\nabla - A)^2 + E]^{-1} \\ &\quad \times [(1-\lambda)^{-1}W]^{1/2} \end{aligned}$$

is the Birman-Schwinger kernel for the operators $\tilde{H} = \tfrac{1}{3}[(i\nabla - A)^2 - 3(1-\lambda)^{-1}W]$. The Birman-Schwinger principle implies that $\int_0^\infty N_1(\tilde{K}_E)dE$ is the sum of the negative eigenvalues of $\tilde{H}$. An estimate on this quantity follows from the standard Lieb-Thirring inequality, i.e.,

$$\begin{aligned} \int_0^\infty N_{1-\lambda}(K_E^>)dE &\leq (0.24)^{\frac{1}{3}} \int \left[\frac{3}{(1-\lambda)} W(x) \right]^2 dx \\ &\leq \tfrac{3}{4}(1-\lambda)^{-2} \int W(x)dx. \end{aligned}$$

The constant 0.24 can be found as $L_{1,2}$ in Ref. 40, Eq. (51). It was improved slightly by Blanchard and Stubbe,⁴¹ see also Ref. 42. In these references only the case $A=0$ was considered. It is, however, a simple consequence of the diamagnetic inequality (see Ref. 43) that the constant is independent of A . Q.E.D.

The Lieb-Thirring inequality in Theorem 5.1 implies an estimate on the kinetic energy

$$T_\psi = \left\langle \psi \left| \sum_{j=1}^N H_{\text{kin}}^{(j)} \right| \psi \right\rangle$$

in terms of the one-particle density

$$\rho_\psi(x) = N \sum_{\sigma_1 = \pm 1/2} \dots \sum_{\sigma_N = \pm 1/2} \int_{\mathbb{R}^{2(N-1)}} |\psi(x, x_2, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 dx_2 \dots dx_N.$$

Here ψ is a normalized N -particle fermionic wave function.

5.2 COROLLARY (kinetic-energy inequality in two dimensions). Let T_ψ and ρ_ψ be defined as above. Then for all $0 < \lambda < 1$ we have

$$T_\psi \geq \begin{cases} 0 & \text{if } \rho_\psi \leq \lambda^{-1} \frac{B}{\pi} \\ \frac{1}{3}(1-\lambda)^2 \int \left[\rho_\psi(x) - \lambda^{-1} \frac{B}{\pi} \right]^2 dx & \text{if } \rho_\psi \geq \lambda^{-1} \frac{B}{\pi}. \end{cases} \quad (5.22)$$

Proof. The inequality in Theorem 5.1 holds for the operator $H_{\text{kin}} - W$ if $|\gamma| \geq 1$. If $|\gamma| < 1$, however, one should

choose Π_0 in the proof of Theorem 5.1 as the projection onto the levels $\nu=0$ and $\nu=1$ (not only onto $\nu=0$). Equation (5.21) is then no longer an identity but a bound. In this way one concludes that the negative eigenvalues $e_1(W)$, $e_2(W)$, ... for $H_{\text{kin}} - W$ satisfy

$$\sum_j |e_j(W)| \leq \alpha \int_{\mathbb{R}^2} |W|_+(x) dx + \beta \int_{\mathbb{R}^2} |W|_+^2(x) dx ,$$

with $\alpha = \lambda^{-1} B / \pi$ and $\beta = \frac{3}{4}(1-\lambda)^{-2}$. This bound is clearly valid for all γ .

The proof of (5.22) now follows by a standard Legendre transformation. In fact, if $W \geq 0$ we have

$$T_\psi = \left\langle \psi \left| \sum_j H_{\text{kin}}^{(j)} - W(x_j) \right| \psi \right\rangle + \int_{\mathbb{R}^2} W \rho_\psi \geq \int_{\mathbb{R}^2} [W \rho_\psi - \alpha W - \beta W^2] . \quad (5.23)$$

Since the Legendre transformation of the function $W \mapsto \alpha W + \beta W^2$ is the function

$$\rho \mapsto \sup_{W \geq 0} [\rho W - \alpha W - \beta W^2] = \begin{cases} 0 & \text{if } \rho \leq \alpha \\ (4\beta)^{-1}(\rho - \alpha)^2 & \text{if } \rho > \alpha , \end{cases}$$

we see that (5.22) follows by making the optimal choice for W in (5.23). Q.E.D.

5.3 LEMMA (exchange inequality in two dimensions). Let $\psi \in \otimes^N L^2(\mathbb{R}^2; \mathbb{C}^2)$ be any normalized N -particle wave function (not necessarily fermionic) and let

$$\rho_\psi(x) = \sum_{i=1}^N \sum_{\sigma_1 = \pm 1/2} \cdots \sum_{\sigma_N = \pm 1/2} \int_{\mathbb{R}^{2N}} |\psi(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N$$

be the corresponding one-particle density. Then

$$\sum_{\sigma_1} \cdots \sum_{\sigma_N} \int_{\mathbb{R}^{2N}} |\psi|^2 \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} dx_1 \cdots dx_N \geq \frac{1}{2} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \rho_\psi(x) \rho_\psi(y) |x - y|^{-1} dx dy - 192(2\pi)^{1/2} \int_{\mathbb{R}^2} \rho_\psi(x)^{3/2} dx . \quad (5.24)$$

Proof. The proof is essentially the same as in Ref. 34, where the three-dimensional equivalent of (1.1) was proved. Our presentation is inspired by Ref. 44.

We use the representation (in three dimensions a similar representation was originally used by Fefferman and de La Llave⁴⁵)

$$|x - y|^{-1} = \pi^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}} \chi_R(x - z) \chi_R(y - z) R^{-4} dz , \quad (5.25)$$

where χ_R is the characteristic function of the ball of radius R centered at the origin. If we use (5.25) to represent $\sum_{i < j} |x_i - x_j|^{-1}$ we can estimate the integrand as follows:

$$\begin{aligned} \sum_{1 \leq i < j \leq N} \chi_R(x_i - z) \chi_R(x_j - z) &= \frac{1}{2} \left[\sum_i \chi_R(x_i - z) \right]^2 - \frac{1}{2} \sum_i \chi_R(x_i - z) \\ &= \frac{1}{2} \left[\sum_i \chi_R(x_i - z) - \int \rho_\psi(y) \chi_R(y - z) dy \right]^2 \\ &\quad + \sum_i \chi_R(x_i - z) \int \rho_\psi(y) \chi_R(y - z) dy - \frac{1}{2} \left[\int \rho_\psi(y) \chi_R(y - z) dy \right]^2 - \frac{1}{2} \sum_i \chi_R(x_i - z) \\ &\geq \sum_i \chi_R(x_i - z) \int \rho_\psi(y) \chi_R(y - z) dy - \frac{1}{2} \left[\int \rho_\psi(y) \chi_R(y - z) dy \right]^2 - \frac{1}{2} \sum_i \chi_R(x_i - z) . \end{aligned}$$

If we integrate this inequality over the measure $R^{-4} dR dz$, the last term $\frac{1}{2} \sum_i \chi_R(x_i - z)$ will give a divergent integral. For the purpose of a lower bound, however, we can restrict the integration in (5.25) to $R > r(z)$, where $r(z) > 0$ is some specific function we shall choose below. Using the fact that

$$\sum_{\sigma_1} \cdots \sum_{\sigma_N} \int_{\mathbb{R}^{2N}} |\psi|^2 \sum_i \chi_R(x_i - z) dx_1 \cdots dx_N = \int_{\mathbb{R}^2} \rho_\psi(y) \chi_R(y - z) dy ,$$

we obtain

$$\begin{aligned}
\sum_1 \cdots \sum_N \int_{\mathbb{R}^{2N}} |\psi|^2 \sum_{i < j} |x_i - x_j|^{-1} dx_1 \cdots dx_N &\geq \frac{1}{2} \pi^{-1} \int_{\mathbb{R}^2} \int_{R > r(z)} \left[\int \rho_\psi(y) \chi_R(y-z) dy \right]^2 R^{-4} dR dz \\
&\quad - \frac{1}{2} \pi^{-1} \int_{\mathbb{R}^2} \int_{R > r(z)} \int \rho_\psi(y) \chi_R(y-z) dy R^{-4} dR dz \\
&\geq \frac{1}{2} \pi^{-1} \int \int \rho_\psi(x) \rho_\psi(y) |x-y|^{-1} dx dy \\
&\quad - \frac{1}{2} \pi^{-1} \int_{\mathbb{R}^2} \int_{R < r(z)} \pi^2 \rho_\psi^*(z)^2 dR dz \\
&\quad - \frac{1}{2} \pi^{-1} \int_{\mathbb{R}^2} \int_{R > r(z)} R^{-2} \pi \rho_\psi^*(z) dR dz. \tag{5.26}
\end{aligned}$$

Here we have introduced the Hardy-Littlewood maximal function

$$\rho_\psi^*(z) = \sup_R (\pi R^2)^{-1} \int \rho_\psi(y) \chi_R(y-z) dy,$$

which, viewed as a map from $L^p(\mathbb{R}^2)$ to $L^p(\mathbb{R}^2)$, is a bounded map for all $p > 1$ (see Ref. 46, pp. 54–58). The error terms in (5.26) can be computed as

$$\frac{1}{2} \pi^{-1} \int_{\mathbb{R}^2} \left[\int_{R < r(z)} \pi^2 \rho_\psi^*(z)^2 dR + \int_{R > r(z)} R^{-2} \pi \rho_\psi^*(z) dR \right] dz = \frac{1}{2} \pi^{-1} \int_{\mathbb{R}^2} [r(z) \pi^2 \rho_\psi^*(z)^2 + r(z)^{-1} \pi \rho_\psi^*(z)] dz.$$

The optimal choice for $r(z)$ is $r(z) = [\pi \rho_\psi^*(z)]^{-1/2}$. This means that the error is $\pi^{1/2} \int_{\mathbb{R}^2} \rho_\psi^*(z)^{3/2} dz$, but this can be estimated by the maximal inequality to be less than $192(2\pi)^{1/2} \int_{\mathbb{R}^2} \rho(z)^{3/2} dz$. Q.E.D.

B. The lower bound

Our goal here is to give a lower bound to $E^Q(N, B, V)$ in terms of $E^{\text{MTF}}(N, B, V)$ with errors of lower order than N^2 as N tends to infinity. It is important here that $V = Nv$, where v is fixed. To be more precise we shall prove that

$$N^{-2} [E^Q(N, B, V) - E^{\text{MTF}}(N, B, V)] \geq -\varepsilon_N^-(v), \tag{5.27}$$

where $\varepsilon_N^-(v)$ is a non-negative function which tends to 0 as $N \rightarrow \infty$ for fixed v . Note, however, that $\varepsilon_N^-(v)$ does not depend on B .

We shall treat the cases of large B and small B separately. In the large- B regime we prove (5.27) by a comparison with the classical models discussed in Secs. III and IV. In the small- B regime we use magnetic coherent states, Theorem 5.1 and Lemma 5.3.

The dividing line between large and small B is determined as follows. If the minimizer $\rho_{1,v}^C$ of $\mathcal{E}^C$, with $\int \rho_{1,v}^C = 1$ and confining potential v , is bounded [e.g., for $v(x) = |x|^2$], then we define small B to mean

$$B/N \leq \beta_c := 2\pi \sup_x \rho_{1,v}^C(x). \tag{5.28}$$

As explained in (3.20) we have for $\beta \geq \beta_c$ that $E^{\text{MTF}}(1, \beta, v) = E^C(1, v)$.

For the general class of v where we do not know the minimizer $\rho_{1,v}^C$ is bounded we simply define

$$\beta_c = N^{1/3}.$$

By Theorem 3.4 we then have that the function

$$\delta(N, v) = \sup_{\beta \geq \beta_c} |E^{\text{MTF}}(1, \beta, v) - E^C(1, v)| \tag{5.29}$$

tends to zero as N tends to infinity.

Case 1, $B/N \geq \beta_c$: By simply ignoring the kinetic-energy operator, which we had normalized to be positive, we have the obvious inequality $E^Q(N, B, V) \geq E^P(N, V)$ where E^P is the energy of the classical point problem.

From Lemma 4.3 we can therefore conclude that

$$E^Q(N, B, V) \geq E^P(N, V) \geq E^C(N, V) - b(v)N^{3/2}. \tag{5.30}$$

Since $E^C(N, V) = N^2 E^C(1, v)$ and $E^{\text{MTF}}(N, B, V) = N^2 E^{\text{MTF}}(1, B/N, v)$ we have from (5.29) that

$$E^Q(N, B, V) \geq E^{\text{MTF}}(N, B, V) - \delta(N, v)N^2 - b(v)N^{3/2}. \tag{5.31}$$

Thus (5.27) holds with $\varepsilon_N^-(v) = \delta(N, v) + b(v)N^{-1/2}$.

We emphasize again that if $\rho_{1,v}^C$ is bounded (e.g., for $v = k|x|^2$) then $\delta(N, v)$ is not needed.

Case 2, $B/N \leq \beta_c$: In this case we use inequality (5.24) to reduce the many-body problem to a one-body problem.

Let ψ be the many-body ground state for H_N . (Since the exterior potential V tends to infinity at infinity H_N will have a ground state.) The correlation estimate (5.24) gives

$$\begin{aligned}
E^Q(N, B, V) &= \langle \psi | H_N | \psi \rangle \\
&\geq \sum_{j=1}^N \langle \psi | H_{\text{kin}}^{(j)} + V(x_j) | \psi \rangle + D(\rho_\psi, \rho_\psi) \\
&\quad - C \int_{\mathbb{R}^2} \rho_\psi^{3/2}(x) dx. \tag{5.32}
\end{aligned}$$

We first estimate the last term in (5.32) in terms of the kinetic energy

$$T_\psi = \left\langle \psi \left| \sum_{j=1}^N H_{\text{kin}}^{(j)} \right| \psi \right\rangle$$

of ψ . According to (5.22) we have

$$\begin{aligned} \int \rho_\psi^{3/2} &\leq (\text{const}) B^{1/2} \int \rho_{\psi < (\text{const})/B} \rho_\psi \\ &+ (\text{const}) \left[\int \rho_\psi \right]^{1/2} \left[\int \rho_{\psi \geq (\text{const})/B} \rho_\psi^2 \right]^{1/2} \\ &\leq (\text{const}) (\sqrt{\beta_c} N^{3/2} + \sqrt{T_\psi} N^{1/2}). \end{aligned} \quad (5.33)$$

Hence, for all $0 < \varepsilon < 1$ (we shall later choose $\varepsilon \sim N^{-1/2}$) we have

$$\begin{aligned} E^Q(N, B, V) &\geq \sum_{j=1}^N \langle \psi | (1 - \varepsilon) H_{\text{kin}}^{(j)} + V(x_j) | \psi \rangle \\ &+ D(\rho_\psi, \rho_\psi) - (\text{const}) (\sqrt{\beta_c} N^{3/2} + \varepsilon^{-1} N), \end{aligned} \quad (5.34)$$

where we have used the fact that

$$\varepsilon T_\psi - \text{const} \sqrt{T_\psi} N^{1/2} \geq -(\text{const}) \varepsilon^{-1} N. \quad (5.35)$$

To relate (5.34) to the MTF problem we use the inequality

$$\begin{aligned} 0 &\leq D(\rho_\psi - \rho^{\text{MTF}}, \rho_\psi - \rho^{\text{MTF}}) \\ &= - \left\langle \psi \left| \sum_{j=1}^N \rho^{\text{MTF}} * |x_j|^{-1} \right| \psi \right\rangle + D(\rho_\psi, \rho_\psi) \\ &+ D(\rho^{\text{MTF}}, \rho^{\text{MTF}}), \end{aligned} \quad (5.36)$$

which is a consequence of the positive definiteness of the kernel $|x - y|^{-1}$. Inserting this in (5.34) gives

$$\begin{aligned} E^Q(N, B, V) &\geq \sum_{j=1}^N \langle \psi | (1 - \varepsilon) H_{\text{kin}}^{(j)} + V(x_j) \\ &+ \rho^{\text{MTF}} * |x_j|^{-1} | \psi \rangle - D(\rho^{\text{MTF}}, \rho^{\text{MTF}}) \\ &- (\text{const}) (\sqrt{\beta_c} N^{3/2} + \varepsilon^{-1} N). \end{aligned} \quad (5.37)$$

Since we have normalized the potential to be positive we have that $(1 - \varepsilon)^{-1} V(x) \geq V(x)$ and also $(1 - \varepsilon)^{-1} \rho^{\text{MTF}} * |x|^{-1} \geq \rho^{\text{MTF}} * |x|^{-1}$.

Hence

$$\begin{aligned} \sum_{j=1}^N e_j &= \sum_{j=1}^N \langle f_j | H_1^{\text{MTF}} | f_j \rangle \geq \sum_{j=1}^N \langle f_j | \tilde{H}_1^{\text{MTF}} | f_j \rangle \\ &= \sum_v \int [\varepsilon_v(B) + \tilde{V}^{\text{MTF}}(u)] \sum_{j=1}^N \langle f_j | \Pi_{vu} | f_j \rangle du - r^{-2} N \int (\nabla g)^2 \\ &+ \sum_{j=1}^N \langle f_j | \tilde{V}^{\text{MTF}} - \tilde{V}^{\text{MTF}} * g_r^2 | f_j \rangle. \end{aligned} \quad (5.41)$$

We first consider the last term. Writing $\sum_{j=1}^N |f_j(x)|^2 = \tilde{\rho}(x)$ we have

$$\begin{aligned} \sum_{j=1}^N \langle f_j | \tilde{V}^{\text{MTF}} - \tilde{V}^{\text{MTF}} * g_r^2(x) | f_j \rangle &= \int_{|x| < R_v + r} [\tilde{V}^{\text{MTF}}(x) - \tilde{V}^{\text{MTF}} * g_r^2(x)] \tilde{\rho}(x) dx \\ &\geq \int_{|x| \leq R_v - r} [V^{\text{MTF}}(x) - V^{\text{MTF}} * g_r^2(x)] \tilde{\rho}(x) dx \\ &- \int_{R_v - r \leq |x| \leq R_v + r} \tilde{V}^{\text{MTF}} * g_r^2(x) \tilde{\rho}(x) dx. \end{aligned}$$

Since $V^{\text{MTF}} = N v^{\text{MTF}}$ and $\int \tilde{\rho} = N$ we have

$$\sum_{j=1}^N \langle f_j | \tilde{V}^{\text{MTF}} - \tilde{V}^{\text{MTF}} * g_r^2 | f_j \rangle \geq -N^2 \sup_{|x| < R_v} |v^{\text{MTF}}(x) - v^{\text{MTF}} * g_r^2(x)| - \left[\sup_{|x| < R_v + r} v^{\text{MTF}}(x) \right] R_v^2 N^2. \quad (5.42)$$

$$\begin{aligned} E^Q(N, B, V) &\geq (1 - \varepsilon) \left\langle \psi \left| \sum_{j=1}^N [H_{\text{kin}}^{(j)} + V(x_j)] \right. \right. \\ &\quad \left. \left. + \rho^{\text{MTF}} * |x_j|^{-1} \right| \psi \right\rangle \\ &- D(\rho^{\text{MTF}}, \rho^{\text{MTF}}) \\ &- (\text{const}) (\sqrt{\beta_c} N^{3/2} + \varepsilon^{-1} N). \end{aligned} \quad (5.38)$$

Obviously

$$\left\langle \psi \left| \sum_{j=1}^N [H_{\text{kin}}^{(j)} + V(x_j)] + \rho^{\text{MTF}} * |x_j|^{-1} \right| \psi \right\rangle \geq \sum_{j=1}^N e_j, \quad (5.39)$$

where $e_1, e_2, \dots, e_N$ are the N lowest eigenvalues of the one-particle Hamiltonian

$$H_1^{\text{MTF}} = H_{\text{kin}} + V(x) + \rho^{\text{MTF}} * |x|^{-1} = H_{\text{kin}} + V^{\text{MTF}}(x). \quad (5.40)$$

We shall estimate $\sum_{j=1}^N e_j$ by a straightforward coherent states analysis.

Let $f_1, \dots, f_N$ be the N lowest normalized eigenfunctions of H_1^{MTF} . For technical reasons we introduce a modified operator $\tilde{H}_1^{\text{MTF}}$ which is obtained from H_1^{MTF} by replacing V^{MTF} by the truncated potential

$$\tilde{V}^{\text{MTF}}(x) = \begin{cases} V^{\text{MTF}}(x), & |x| \leq R_v \\ CN, & |x| \geq R_v, \end{cases}$$

where R_v is the finite radius given in the Appendix and $C = \inf_{|x| > R_v} V^{\text{MTF}}(x)/N$ is independent of N by the scaling (2.9) of MTF theory. Note that $V^{\text{MTF}} \geq \tilde{V}^{\text{MTF}}$. Then from (5.11) and (5.12) we have

We can then write (5.41) as

$$\sum_{j=1}^N e_j \geq \sum_v \int [\varepsilon_v(B) + \tilde{V}^{\text{MTF}}(u)] \times \sum_{j=1}^N \langle f_j | \Pi_{vu} | f_j \rangle du - N^2 \tilde{\varepsilon}_N(v), \quad (5.43)$$

where

$$\tilde{\varepsilon}_N(v) = \sup_{|x| < R_v} |v^{\text{MTF}}(x) - v^{\text{MTF}} * g_r^2(x)| + \left[\sup_{|x| < R_v + r} v^{\text{MTF}}(x) \right] R_v^2 r - N^{-1} r^2 \int (\nabla g)^2 \leq C(v) N^{-1/3}. \quad (5.44)$$

For the last step we made the choice $r \sim N^{-1/3}$.

We focus next on the first term in (5.43). It has the form

$$\int [\varepsilon_v(B) + \tilde{V}^{\text{MTF}}(u)] \rho_v(u) du, \quad (5.45)$$

where we have denoted $\sum_{j=1}^N \langle f_j | \Pi_{vu} | f_j \rangle$ by $\rho_v(u)$. These functions satisfy

$$0 \leq \rho_v(u) \leq \text{Tr} \Pi_{vu} = d_v(B) \quad (5.46)$$

and

$$\sum_v \int \rho_v(u) du = N. \quad (5.47)$$

We obtain a lower bound to (5.45) by minimizing over all functions ρ_v satisfying (5.46) and (5.47).

Minimizers ρ_v can be constructed as follows. There is a $\mu > 0$ such that

$$\rho_v(u) = \begin{cases} d_v(B) & \text{if } \varepsilon_v(B) + \tilde{V}^{\text{MTF}}(u) < \mu \\ 0 & \text{if } \varepsilon_v(B) + \tilde{V}^{\text{MTF}}(u) > \mu \\ \leq d_v(B) & \text{if } \varepsilon_v(B) + \tilde{V}^{\text{MTF}}(u) = \mu. \end{cases} \quad (5.48)$$

All families ρ_v satisfying (5.48) and the constraint (5.47) are minimizers. Note that it is possible that $\varepsilon_v(B) + \tilde{V}^{\text{MTF}}(u) = \mu$ on an open set of u values. The minimizers are therefore not necessarily unique. The chemical potential μ is uniquely determined by (5.48) and the condition (5.47).

We shall now prove that $\mu = \mu^{\text{MTF}}$. All we have to show is that we can find functions ρ_v satisfying (5.47) and (5.48) with $\mu = \mu^{\text{MTF}}$.

We know from the MTF equation, Theorem 2.2, that if $\rho^{\text{MTF}}(u) = 0$ then $V^{\text{MTF}}(u) \geq \mu^{\text{MTF}}$. Since $\tilde{V}^{\text{MTF}}$ differs from V^{MTF} only on the set $\rho^{\text{MTF}} = 0$ we may in (5.48) when $\mu = \mu^{\text{MTF}}$ replace $\tilde{V}^{\text{MTF}}$ by V^{MTF} . We then know from the MTF equation that there are *unique* functions ρ_v satisfying (5.48) with $\mu = \mu^{\text{MTF}}$ and $\sum_v \rho_v(u) = \rho^{\text{MTF}}(u)$. In fact, in terms of the filling factors (5.17) we have $\rho_v(u) = f_v(u) d_v(B)$. Since $\int \rho^{\text{MTF}}(u) du = N$ we have produced the functions ρ_v allowing us to conclude that μ is indeed equal to μ^{MTF} . If we insert these functions in (5.45) we obtain

$$\sum_v \int \varepsilon_v(B) \rho_v(u) du + \int V^{\text{MTF}}(u) \rho^{\text{MTF}}(u) du = \int \{j_B[\rho^{\text{MTF}}(u)] + V^{\text{MTF}}(u)\} du, \quad (5.49)$$

where the identity follows from (2.3).

We can now combine (5.38), (5.39), (5.43), (5.44), and (5.49) to arrive at

$$E^Q(N, B, V) \geq (1 - \varepsilon) E^{\text{MTF}}(N, B, V) - \varepsilon D(\rho^{\text{MTF}}, \rho^{\text{MTF}}) - (\text{const})(\sqrt{\beta_c} N^{3/2} + \varepsilon^{-1} N) - C(v) N^{5/3}. \quad (5.50)$$

Hence, since $D(\rho^{\text{MTF}}, \rho^{\text{MTF}}) \leq E^{\text{MTF}}(N, B, V)$ we have

$$N^{-2} [E^Q(N, B, V) - E^{\text{MTF}}(N, B, V)] \geq -2\varepsilon E^{\text{MTF}}(1, B/N, v) - (\text{const})[\sqrt{\beta_c} N^{-1/2} + \varepsilon^{-1} N^{-1} + c(v) N^{-1/3}].$$

Note that $E^{\text{MTF}}(1, B/N, v)$ is bounded by a constant depending only on v . If we choose $\varepsilon \sim N^{-1/2}$ we find

$$N^{-2} [E^Q(N, B, V) - E^{\text{MTF}}(N, B, V)] \geq -c^-(v) [\sqrt{\beta_c} N^{-1/2} + N^{-1/3}].$$

This is equivalent to (5.27) with $\varepsilon_N^-(v) = c^-(v) [\sqrt{\beta_c} N^{-1/2} + N^{-1/3}]$. In the case when ρ^C is bounded, β_c is a constant, otherwise we chose it to be $\beta_c = N^{1/3}$. In both cases $\varepsilon_N^-(v)$ will tend to zero as N tends to infinity. This finishes the proof of the lower bound.

We have proved (1.7). The limits in (1.8) and (1.9) follow immediately from the corresponding results for E^{MTF} proved in Secs. II and III.

C. Homogeneous exterior potentials

Finally, we shall show how to prove the stronger result Theorem 1.3 for homogeneous exterior potentials.

In this case we do not have that the minimizing MTF density is supported within a fixed ball. In fact, the density will spread out as the coupling constant becomes small.

We shall prove that given $\varepsilon > 0$ and k_0 there is an N_ε independent of B such that for $N \geq N_\varepsilon$ and $K/N \leq k_0$

$$|E^Q(N, B, Kv)/E^{\text{MTF}}(N, B, Kv) - 1| < \varepsilon. \quad (5.51)$$

We consider large and small K in very much the same way as we did for B in the lower bound above. We shall see below that we can find a k_c (depending only on v and ε) such that (5.51) holds for $K/N \leq k_c$.

In the case of large K , i.e., $K/N \geq k_c$ (but $K/N \leq k_0$) the proof of (5.51) is then identical to the proof of Theorem 1.1 given above.

For small K we again consider the upper and lower bounds separately. We begin with the upper bound. We

proceed as in Sec. V A. We define the trial operator as in (5.17) and (5.18) except that we replace ρ^{MTF} by $\rho_{N,Kv}^C$. The estimate (5.20) now becomes

$$\begin{aligned} E^Q(N, B, Kv) &\leq \mathcal{E}^{\text{MTF}}[\rho_{N,Kv}^C; B, Kv] \\ &\quad + Nr^{-2} \int [\nabla g(x)]^2 dx \\ &\quad + NK \sup_{|x| \leq R_K} [v * g_r^2(x) - v(x)]. \end{aligned}$$

Here R_K is the radius of the ball containing the support

of $\rho_{N,Kv}^C$ for $K = kN$. According to (3.22) $R_k = k^{-1/(s+1)}R_1$, where R_1 is the radius for $k=1$, which depends only on v .

Using the homogeneity of v we have

$$\begin{aligned} v * g_r^2(x) - v(x) &= |x|^s \int \{v((x-y)/|x|) \\ &\quad - v(x/|x|)\} g_r^2(y) dy \\ &\leq c_3(v)r|x|^{s-1}. \end{aligned}$$

We obtain since $s \geq 1$

$$\begin{aligned} E^Q(N, B, Kv) - \mathcal{E}^{\text{MTF}}[\rho_{N,Kv}^C; B, Kv] &\leq c_4(v)N[r^{-2} + KrR^{s-1}] \\ &\leq c_5(v)N[r^{-2} + Kr(K/N)^{-(s-1)/(s+1)}] \\ &\leq c_5(v)N^{2/3}(K/N)^{4/[3(s+1)]}, \end{aligned}$$

with the choice $r = (K/N)^{-2/[3(s+1)]}$. We also know from (3.25) that

$$\mathcal{E}^{\text{MTF}}[\rho_{N,Kv}^C; B, Kv] \leq (K/N)^{2/(s+1)}N^2 \int j_0(\rho_{1,v}^C) + E^C(N, Kv).$$

Hence, from (3.21) and (3.17) we obtain

$$\begin{aligned} E^Q(N, B, Kv) &\leq \left[1 + \frac{c_6(v)N^2}{E^C(N, Kv)} [N^{-1/3}(K/N)^{4/[3(s+1)]} + (K/N)^{2/(s+1)}] \right] E^C(N, Kv) \\ &\leq (1 + c_7(v)[N^{-1/3}(K/N)^{1/[3(s+1)]} + (K/N)^{1/(s+1)}]) E^C(N, Kv) \end{aligned}$$

It therefore follows from Theorem 3.5 that we can find k_ε depending only on ε (but not on B) such that $E^Q(N, B, Kv)/E^{\text{MTF}}(N, B, Kv) \leq 1 + \varepsilon$ for $K/N \leq k_\varepsilon$.

We turn next to the lower bound. As in Sec. V B (in the case $B/N \geq \beta_c$) we may ignore the kinetic-energy operator, which we had normalized to be positive. We then have the obvious inequality $E^Q(N, B, V) \geq E^P(N, V)$ where E^P is the energy of the classical point problem.

We shall use Lemma 4.3 to compare $E^P(N, Kv)$ to $E^C(N, Kv)$. We must, however, first discuss the scaling of $E^P(N, Kv)$. It is clear that if v is homogeneous of degree s then

$$\begin{aligned} \mathcal{E}^P(x_1, \dots, x_N; Kv) \\ = k^{1/(s+1)} \mathcal{E}^P(k^{1/(s+1)}x_1, \dots, k^{1/(s+1)}x_N; v). \end{aligned}$$

Therefore, $E^P(N, kV) = k^{1/(s+1)}E^P(N, V)$, i.e., $E^P(N, kV)$ has the same scaling as $E^C(N, Kv)$ [see (3.21)].

From Lemma 4.3 we thus find that

$$\begin{aligned} E^P(N, Kv) &\geq E^C(N, Kv) - b(v)(K/N)^{1/(s+1)}N^{3/2} \\ &\geq E^C(N, Kv)[1 - c_8(v)N^{-1/2}]. \end{aligned} \quad (5.52)$$

According to Theorem 3.5 we may thus assume that k_ε is such that

$$\begin{aligned} E^Q(N, B, Kv)/E^{\text{MTF}}(N, B, Kv) \\ \geq (1 - \varepsilon/2)[1 - c_8(v)N^{-1/2}] \end{aligned} \quad (5.53)$$

for $K/N \leq k_\varepsilon$. We can therefore clearly find N_ε such that the right-hand side of (5.53) is greater than $1 - \varepsilon$ for $N \geq N_\varepsilon$.

VI. CONCLUSION

We have analyzed the ground state of a two-dimensional gas of N electrons interacting with each other via the (three-dimensional) Coulomb potential and subject to a confining exterior potential $V(x) = Kv(x)$ where K is an adjustable coupling constant. The electrons are also subject to a uniform magnetic field B perpendicular to the two-dimensional plane.

We have found the exact energy and electron density function $\rho(x)$ to leading order in $1/N$, i.e., in the high-density limit. This limit is achieved by letting K be proportional to N as $N \rightarrow \infty$, thus effectively confining the electrons to a fixed region of space, independent of N .

It turns out that the answer to the problem depends critically on the behavior of B as $N \rightarrow \infty$. There are three regimes.

(i) If $B/N \rightarrow 0$, i.e., $N \gg B$ in appropriate units, then normal (two-dimensional) Thomas-Fermi theory gives the exact description. Correlations can be ignored to leading order in this high-density situation.

(ii) If $B/N = \text{const}$, a modified TF theory in which the “kinetic-energy density” is changed from $(\text{const})\rho^2$ to a certain B -dependent function of ρ [called $j_B(\rho)$] is exact.

(iii) If $B/N \rightarrow \infty$ then the kinetic-energy term can be omitted entirely and a classical continuum electrostatics

theory emerges as the exact theory. This electrostatics problem is mathematically interesting in its own right and can be solved in closed form for the customary choice $v(x) = |x|^2$.

Related to the continuum problem is an electrostatics problem for point charged particles. Apart from its mathematical interest, it provides a crucial lower bound to the energy in case (iii). Another technical point of some interest is the extension of the Lieb-Thirring inequality to two-dimensional particles in a magnetic field which involves dealing with a continuum of zero energy modes (i.e., the lowest Landau level).

ACKNOWLEDGMENTS

We thank Vidar Gudmundsson and Jari Kinaret for valuable discussions. We thank Kristinn Johnsen for allowing us to use his graphs of the electron density (Fig. 2). E.H.L. was partially supported by U.S. National Sci-

ence Foundation Grant No. PHY90-19433 A03. J.P.S. was partially supported by U.S. National Science Foundation Grant No. DMS 92-03829. J.Y. was partially supported by the Icelandic Science Foundation and the Research Fund of the University of Iceland.

APPENDIX

Here we prove that the minimizers for our three semi-classical problems can be sought among densities that vanish outside some finite radius—for which we give an upper bound. This lemma is in an appendix because it pertains to several sections of the paper.

A.1 LEMMA (finite radius of minimizers). Consider the three cases: (a) The classical energy; (b) the classical particle energy; (c) the MTF energy. Let $V(x)$ be the confining potential. We assume that $V(x) \rightarrow +\infty$ as $|x| \rightarrow \infty$ in the sense that the number $W(R) := \inf\{V(x) : |x| \geq R\}$ tends to ∞ as $R \rightarrow \infty$.

Then there is a radius R_v , depending only on $v = V/N$ such that

$$\begin{aligned} E^P(N, V) &= \inf\{\mathcal{E}^P(x_1, \dots, x_N) : |x_i| \leq R_v \text{ for all } i\} \\ E^C(N, V) &= \inf\left\{\mathcal{E}^C[\mu; V] : \text{support } \mu \subset \{x : |x| \leq R_v\}, \int d\mu = N\right\} \\ E^{\text{MTF}}(N, B, V) &= \inf\left\{\mathcal{E}^{\text{MTF}}[\rho; B, V] : \rho(x) = 0 \text{ for } |x| > R_v, \int \rho = N\right\}. \end{aligned} \quad (\text{A1})$$

Furthermore, any minimizing particle distribution measure or density satisfies the conditions given in braces in (A1).

A choice for R_v , which is far from optimal, is any R satisfying the inequality

$$\frac{1}{N} W(R) \geq (2 + \pi^{-1}) + \frac{1}{N} \langle V \rangle_1, \quad (\text{A2})$$

with $\langle V \rangle_1$ being the average of V in the unit disk:

$$\langle V \rangle_1 = \frac{1}{\pi} \int_{|x| < 1} V(x) dx.$$

Proof: Particle case. Suppose that $|x_1| > R_v$. Then we move particle 1 inside D , the unit disc centered at the origin. The point y to which we move particle 1 is not known, so we average the energy over all choices of $y \in D$. If we show that this average energy is less than the original energy then we know that there is some point $y \in D$ such that the energy is lowered. Thus, we have to show that

$$V(x_1) + \sum_{j=2}^N |x_1 - x_j|^{-1} > \langle V \rangle_1 + \frac{1}{\pi} \sum_{j=2}^N \int_D |y - x_j|^{-1} dy.$$

Noting that $\int_D |y - x|^{-1} dy \leq \int_D |y|^{-1} dy = 2\pi$, by a simple rearrangement inequality, we see that it suffices to have $W(R_v) > \langle V \rangle_1 + 2N$, which agrees with (A2).

The classical case. If μ is any measure with $\int d\mu = N$, we define μ^+ to be μ restricted to the complement of the closed disc of radius R_v centered at the origin. Thus $\mu^+(A) = \mu(A \cap \{x : |x| > R_v\})$. Similarly, μ^- is μ restricted to the disc, so that $\mu = \mu^+ + \mu^-$. Assuming that $\mu^+ \neq 0$, we replace μ by $\mu_\epsilon := (1 - \epsilon)\mu^+ + \mu^- + \delta\nu$, where ν is Lebesgue measure restricted to the unit disc D , and where $\pi\delta = \epsilon \int d\mu^+$. Thus $\int d\mu_\epsilon = N$. The change in energy, to $O(\epsilon)$ as $\epsilon \downarrow 0$, is easily seen to be

$$\begin{aligned} \delta \int_D V(x) dx - \epsilon \int V(x) \mu^+(dx) + \delta \int_D \int_{\mathbb{R}^2} |x - y|^{-1} dx \mu(dy) \\ - \epsilon \int \int |x - y|^{-1} \mu^+(dx) \mu(dy) < \pi\delta \langle V \rangle_1 - \epsilon W(R_v) \int d\mu^+ + 2\pi\delta N, \end{aligned}$$

which is negative by (A1).

The MTF case. This is similar to the classical case, but with two differences: (i) The measure μ is replaced by a function ρ with $\int \rho(x)dx = N$ and (ii) a “kinetic-energy” term $\int j_B[\rho(x)]dx$ is added to the energy. Point (i) only simplifies matters. For point (ii) we note the simple fact that $j_B(\rho)$ is bounded above by $\pi\rho^2/2$ and its derivative, $j'_B(\rho)$, is bounded above by $\pi\rho$; this is true for all B . Let us assume that $d\mu^+ := \rho^+(x)dx$ is not zero, with $\rho^+(x) = \rho(x)$ for $|x| > R_\nu$ and $\rho^+(x) = 0$ otherwise. The argument is as before, but now we must take into account the change in kinetic energy which, to leading order in ε , is

$$\delta \int_D j'_B[\rho(x)]dx - \varepsilon \int \rho^+(x) j'_B[\rho^+(x)]dx < \delta \pi \int_D \rho(x)dx \leq \delta N = \varepsilon \int \rho^+(x)dx.$$

The total energy change is then negative by (A1). Q.E.D.

- ¹H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in *Single Charge Tunneling*, edited by H. Grabert, J. M. Martinis, and M. H. Devoret (Plenum, New York, 1991).
- ²M. A. Kastner, Rev. Mod. Phys. **64**, 849 (1992).
- ³P. L. McEuen, E. B. Foxman, J. Kinaret, U. Meirav, M. A. Kastner, N. S. Wingreen, and S. J. Wind, Phys. Rev. B **45**, 11 419 (1992).
- ⁴R. C. Ashoori, H. L. Stormer, J. S. Weiner, L. N. Pfeiffer, S. J. Pearton, K. W. Baldwin, and K. W. West, Phys. Rev. Lett. **68**, 3088 (1992).
- ⁵R. C. Ashoori, H. L. Stormer, J. S. Weiner, L. N. Pfeiffer, K. W. Baldwin, and K. W. West, Phys. Rev. Lett. **71**, 613 (1993).
- ⁶N. C. van der Vaart, M. P. de Ruyter van Steveninck, L. P. Kouwenhoven, A. T. Johnson, Y. V. Nazarov, and C. J. P. M. Harmans, Phys. Rev. Lett. **73**, 320 (1994).
- ⁷O. Klein, C. Chamon, D. Tang, D. M. Abush-Magder, X.-G. Wen, M. A. Kastner, and S. J. Wind (unpublished).
- ⁸A. Kumar, S. E. Laux, and F. Stern, Phys. Rev. B **42**, 5166 (1990).
- ⁹C. W. J. Beenakker, Phys. Rev. B **44**, 1646 (1991).
- ¹⁰V. Shikin, S. Nazin, D. Heitmann, and T. Demel, Phys. Rev. B **43**, 11 903 (1991).
- ¹¹V. Gudmundsson and R. R. Gerhardt, Phys. Rev. B **43**, 12 098 (1991).
- ¹²A. H. MacDonald, S. R. Eric Yang, and M. D. Johnson, Aust. J. Phys. **46**, 345 (1993).
- ¹³P. L. McEuen, N. S. Wingreen, E. B. Foxman, J. Kinaret, U. Meirav, M. A. Kastner, and S. J. Wind, Physica B **189**, 70 (1993).
- ¹⁴S.-R. Eric Yang, A. H. MacDonald, and M. D. Johnson, Phys. Rev. Lett. **71**, 3194 (1993).
- ¹⁵J. M. Kinaret and N. S. Wingreen, Phys. Rev. B **48**, 11 113 (1993).
- ¹⁶D. Pfannkuche, V. Gudmundsson, and P. A. Maksym, Phys. Rev. B **47**, 2244 (1993).
- ¹⁷D. Pfannkuche, V. Gudmundsson, P. Hawrylak, and R. R. Gerhardt, Solid State Electron. **37**, 1221 (1994).
- ¹⁸M. Ferconi and G. Vignale (unpublished).
- ¹⁹E. H. Lieb, J. P. Solovej, and J. Yngvason, Phys. Rev. Lett. **69**, 749 (1992).
- ²⁰E. H. Lieb, J. P. Solovej, and J. Yngvason, Commun. Pure Appl. Math. **47**, 513 (1994).
- ²¹E. H. Lieb, J. P. Solovej, and J. Yngvason, Commun. Math. Phys. **161**, 77 (1994).
- ²²E. H. Lieb, J. P. Solovej, and J. Yngvason, in *Proceedings of the Conference on Partial Differential Equations and Mathematical Physics, Birmingham, AL, 1994*, edited by I. Knowles (International Press, Cambridge, MA, in press).
- ²³Y. Tomishima and K. Yonei, Progr. Theor. Phys. **59**, 683 (1978).
- ²⁴I. Fushiki, E. H. Gudmundsson, C. J. Pethick, and J. Yngvason, Ann. Phys. (N.Y.) **216**, 29 (1992).
- ²⁵J. Yngvason, Lett. Math. Phys. **22**, 107 (1991).
- ²⁶E. H. Lieb and H.-T. Yau, Commun. Math. Phys. **118**, 177 (1988).
- ²⁷V. Fock, Z. Phys. **47**, 446 (1928).
- ²⁸E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981); **54**, 311(E) (1982).
- ²⁹E. H. Lieb and B. Simon, Adv. Math. **23**, 22 (1977).
- ³⁰E. H. Lieb and M. Loss (unpublished).
- ³¹E. H. Lieb and W. E. Thirring, Phys. Rev. Lett. **35**, 687 (1975).
- ³²E. H. Lieb and W. E. Thirring, in *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*, edited by E. H. Lieb, B. Simon, and A. Wightman (Princeton University Press, Princeton, NJ, 1976), pp. 269–303.
- ³³L. Erdős, Commun. Math. Phys. (to be published).
- ³⁴E. H. Lieb, Phys. Lett. **70A**, 444 (1979).
- ³⁵E. H. Lieb and S. Oxford, Int. J. Quantum Chem. **19**, 427 (1981).
- ³⁶E. H. Lieb and J. P. Solovej, Lett. Math. Phys. **22**, 145 (1991).
- ³⁷E. H. Lieb, Phys. Rev. Lett. **46**, 457 (1981); **47**, 69(E) (1981).
- ³⁸B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).
- ³⁹K. Fan, Proc. Natl. Acad. Sci. U.S.A. **37**, 760 (1951).
- ⁴⁰E. H. Lieb, Commun. Math. Phys. **92**, 57 (1984).
- ⁴¹Ph. Blanchard and J. Stubbe (unpublished).
- ⁴²A. Martin, Commun. Math. Phys. **129**, 161 (1990).
- ⁴³E. H. Lieb, in *Schrödinger Operators*, Proceedings of a Conference in Sønderborg, Denmark, 1988, edited by H. Holden and A. Jensen, Springer Lecture Notes in Physics Vol. 345 (Springer, Berlin, 1989), pp. 371–382.
- ⁴⁴V. Bach, Commun. Math. Phys. **147**, 527 (1992).
- ⁴⁵C. Fefferman and R. de la Llave, Rev. Iberoamericana **2**, 119 (1986).
- ⁴⁶E. M. Stein and G. Weiss, *Fourier Analysis on Euclidean Spaces* (Princeton University Press, Princeton, NJ, 1971).

Part III

General Results

with Applications to Atoms

KINETIC ENERGY BOUNDS AND THEIR APPLICATION TO THE STABILITY OF MATTER

Elliott H. Lieb

Departments of Mathematics and Physics, Princeton University

P.O. Box 708, Princeton, NJ 08544

The Sobolev inequality on $\mathbf{R}^n$, $n \geq 3$ is very important because it gives a lower bound for the kinetic energy $\int |\nabla f|^2$ in terms of an L^p norm of f . It is the following.

$$\int_{\mathbf{R}^n} |\nabla f|^2 \geq S_n \left\{ \int_{\mathbf{R}^n} |f|^{2n/(n-2)} \right\}^{(n-2)/n} = S_n \|f\|_{2n/(n-2)}^2. \quad (1)$$

Applying Hölder's inequality to the right side we obtain the following modification of (1).

$$\int_{\mathbf{R}^n} |\nabla f|^2 \geq K_n^1 \left\{ \int_{\mathbf{R}^n} \rho^{(n+2)/n} \right\} / \left\{ \int_{\mathbf{R}^n} \rho \right\}^{2/n} = K_n^1 \|\rho\|_{(n+2)/n}^{(n+2)/n} \|\rho\|_1^{-2/n} \quad (2)$$

with $\rho(x) \equiv |f(x)|^2$. The superscript 1 on K_n^1 indicates that in (2) we are considering only one function, f . Hölder's inequality implies that $K_n^1 \geq S_n$ but, in fact, the *sharp* value of K_n^1 (which can be obtained by solving a nonlinear PDE) is larger than S_n . In particular, $K_n^1 > 0$ for *all* $n \geq 1$, even though $S_n = 0$ for $n < 3$.

Inequality (2), unlike (1) has the following important property: The non-linear term $\int \rho^{(n+2)/n}$ enters with the power 1 (and not $(n-2)/n$) and is therefore "extensive." The price we have to pay for this is the factor $\|f\|_2^{4/n} = \|\rho\|_1^{2/n}$ in the denominator, but since we shall apply (2) to cases in which $\|f\|_2 = 1$ (L^2 normalization condition) this is not serious.

Inequality (2) is equivalent to the following: Consider the Schrödinger operator on $\mathbf{R}^n$

$$H = -\Delta - V(x) \quad (3)$$

and let $e_1 = \inf \text{spec}(H)$. (We assume H is self-adjoint.) Let $V_+(x) \equiv \max\{V(x), 0\}$. Then

$$e_1 \geq -L_{1,n}^1 \int V_+(x)^{(n+2)/n} dx = -L_{1,n}^1 \|V_+\|_{(n+2)/2}^{(n+2)/2} \quad (4)$$

with

$$L_{1,n}^1 = \left(\frac{n}{2K_n^1} \right)^{n/2} \left(1 + \frac{n}{2} \right)^{-(n+2)/n}. \quad (5)$$

The reason for the subscript 1 in $L_{1,n}^1$ will be clarified in eq. (8).

Here is the proof of the equivalence. We have

$$e_1 \geq \inf_f \left\{ \int |\nabla f|^2 - \int \rho V_+ \mid \|f\|_2 = 1 \text{ and } \rho = |f|^2 \right\}.$$

Use (2) and Hölder to obtain (with $X = \|\rho\|_{(n+2)/n}$)

$$e_1 \geq \inf_X \left\{ K_n^1 X^{(n+2)/n} - \|V_+\|_{(n+2)/2} X \right\} \quad (6)$$

Minimizing (6) with respect to X yields (4). To go from (4) to (2), take $V = V_+ = \alpha|f|^{4/n} = \alpha\rho^{2/n}$ in (3). Then

$$-L_{1,n}^1 \alpha^{(n+2)/2} \int \rho^{(n+2)/n} \leq e_1 \leq (f, Hf) = \int |\nabla f|^2 - \alpha \int \rho^{(n+2)/n}.$$

Optimizing this with respect to α yields (2).

So far this is trivial, but now we turn to a more interesting question. Let $e_1 \leq e_2 \leq \dots \leq 0$ be the negative spectrum of H (which may be empty). Is there a bound of the form

$$\sum e_i \geq -L_{1,n} \int V_+(x)^{(n+2)/2} dx \quad (7)$$

for some universal, V independent, constant $L_{1,n} > 0$ (which, of course, is $\geq L_{1,n}^1$)? The point is that the right side of (7) has the same form as the right side of (4). More generally, given $\gamma \geq 0$, does

$$\sum |e_i|^\gamma \leq L_{\gamma,n} \int V_+(x)^{\gamma + \frac{n}{2}} \quad (8)$$

hold for suitable $L_{\gamma,n}$? When $\gamma = 0$, $\sum |e_i|^0$ is interpreted as the number of $e_i \leq 0$.

The answer to these questions is yes in the following cases:

$n = 1$: All $\gamma > \frac{1}{2}$. The case $\gamma = 1/2$ is unsettled. For $\gamma < \frac{1}{2}$, examples show there can be no bound of the form (8).

$n = 2$: All $\gamma > 0$. There can be no bound when $\gamma = 0$.

$n \geq 3$: All $\gamma \geq 0$.

The cases $\gamma > 0$ were first done in [10], [11]. The $\gamma = 0$ case for $n \geq 3$ was done in [3], [6], [14], with [6] giving the best estimate for $L_{0,n}$. For a review of what is currently known about these constants and conjectures about the sharp values of $L_{\gamma,n}$, see [8]. The proof of (8) is involved (especially when $\gamma = 0$) and will not be given here. It uses the Birman-Schwinger kernel, $V_+^{1/2}(-\Delta + \lambda)^{-1}V_+^{1/2}$.

There is a natural “guess” for $L_{\gamma,n}$ in terms of a semiclassical approximation (and which is not unrelated to the theory of pseudodifferential operators):

$$\sum |e_i|^\gamma \approx (2\pi)^{-n} \int_{\mathbf{R}^n \times \mathbf{R}^n, p^2 \leq V(x)} [V(x) - p^2]^\gamma dp dx \quad (9)$$

$$= L_{\gamma,n}^c \int V_+(x)^{\gamma+n/2} dx. \quad (10)$$

From (9),

$$L_{\gamma,n}^c = (4\pi)^{-n/2} \Gamma(\gamma+1) / \Gamma(1+\gamma+n/2). \quad (11)$$

It is easy to prove that

$$L_{\gamma,n} \geq L_{\gamma,n}^c. \quad (12)$$

The evaluation of the sharp $L_{\gamma,n}$ is an interesting open problem – especially $L_{1,n}$. In particular, for which γ, n is $L_{\gamma,n} = L_{\gamma,n}^c$? It is known [1] that for each fixed n , $L_{\gamma,n}/L_{\gamma,n}^c$ is nonincreasing in γ . Thus, if $L_{\gamma_0,n} = L_{\gamma_0,n}^c$ for some γ_0 , then $L_{\gamma,n} = L_{\gamma,n}^c$ for all $\gamma > \gamma_0$. In particular, $L_{3/2,1} = L_{3/2,1}^c$ [11], so $L_{\gamma,1} = L_{\gamma,1}^c$ for $\gamma \geq 3/2$. No other sharp values of $L_{\gamma,n}$ are known. It is also known [11] that $L_{\gamma,1} > L_{\gamma,1}^c$ for $\gamma < 3/2$ and $L_{\gamma,n} > L_{\gamma,n}^c$ for $n = 2, 3$ and small γ .

Just as (4) is related to (2), inequality (7) is related to a generalization of (2). (The proof is basically the same.) Let $\phi_1, \dots, \phi_N$ be any set of L^2 orthonormal functions on $\mathbf{R}^n (n \geq 1)$ and define

$$\rho(x) \equiv \sum_{i=1}^N |\phi_i(x)|^2. \quad (13)$$

$$T \equiv \sum_{i=1}^N \int |\nabla \phi_i|^2 \quad (14)$$

Then we have **The Main Inequality**

$$T \geq K_n \int \rho(x)^{1+2/n} dx \quad (15)$$

with K_n related to $L_{1,n}$ as in (5), i.e.

$$L_{1,n} = \left(\frac{n}{2K_n} \right)^{n/2} \left(1 + \frac{n}{2} \right)^{-(n+2)/2} \quad (16)$$

The best current value of K_n , for $n = 1, 2, 3$ is in [8]; in particular $K_3 \geq 2.7709$. We might call (15) a *Sobolev type inequality for orthonormal functions*. The point is that if the ϕ_i are merely normalized, but not orthogonal, then the best one could say is

$$T \geq N^{-2/n} K_n^1 \int \rho(x)^{1+2/n} dx. \quad (17)$$

The orthogonality eliminates the factor $N^{-2/n}$, but replaces K_n^1 by the slightly smaller value K_n .

One should notice, especially, the N dependence in (15). The right side, loosely speaking, is proportional to $N^{(n+2)/n}$, whereas the right side of (17) appears, falsely, to be proportional to N^1 , which is the best one could hope for without orthogonality. The difference is crucial for applications. In fact, if one is willing to settle for N^1 one can proceed directly from (1) (for $n \geq 3$). One then has (with $p = n/(n-2)$)

$$T \geq S_n \left\{ \int \rho(x)^p dx \right\}^{1/p}, \quad (n \geq 3). \quad (17a)$$

This follows from $\sum \|\phi_i^2\|_p \geq \|\sum |\phi_i|^2\|_p$.

Eq. (11) gives a “classical guess” for $L_{1,n}$. Using that, together with (16), we have a “classical guess” for K_n , namely

$$\begin{aligned} K_n^c &= 4\pi n \Gamma\left(\frac{n+2}{2}\right)^{2/n} / (2+n) \\ &= \frac{3}{5} (6\pi^2)^{2/3} = 9.1156 \text{ for } n = 3. \end{aligned} \quad (18)$$

Since $L_{1,n} \geq L_{1,n}^c$, we have $K_n \leq K_n^c$. A conjecture in [11] is that $K_3 = K_3^c$, and it would be important to settle this.

Inequality (15) can be easily extended to the following: Let $\psi(x_1, \dots, x_N) \in L^2((\mathbf{R}^n)^N)$, $x_i \in \mathbf{R}^n$. Suppose $\|\psi\|_2 = 1$ and ψ is *antisymmetric* in the N variables, i.e., $\psi(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = -\psi(x_1, \dots, x_j, \dots, x_i, \dots, x_N)$. Define

$$\rho_i(x) = \int |\psi(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_N)|^2 dx_1 \dots \widehat{dx_i} \dots dx_N \quad (19)$$

$$T_i(x) = \int |\nabla_i \psi|^2 dx_1 \dots dx_N \quad (20)$$

$$\rho(x) = \sum_{i=1}^N \rho_i(x) \quad T = \sum_{i=1}^N T_i. \quad (21)$$

(Note that $\rho(x) = N\rho_1(x)$ and $T = NT_1$ since ψ is antisymmetric, but the general form (19)-(21) will be used in the next paragraph.) **Then (15) holds with ρ and T given by (19)-(21) (with the same K_n as in (15)).** This is a generalization of (13)-(15) since we can take

$$\psi(x_1, \dots, x_N) = (N!)^{-1/2} \det \{\phi_i(x_j)\}_{i,j=1}^N,$$

which leads to (13) and (14).

A variant of (15) is given in (52) below. It is a consequence of the fact that (17) and (17a) *also hold* with the definitions (19)-(21). *Antisymmetry of ψ is not required.* The proof of (17a) just uses (1) as before plus Minkowski's inequality, namely for $p \geq 1$

$$\int \left\{ \int |F(x, y)|^p dy \right\}^{1/p} dx \geq \left\{ \int \left\{ \int |F(x, y)| dx \right\}^p dy \right\}^{1/p}.$$

We turn now to some applications of these inequalities.

Application 1. Inequality (15) can be used to bound L^p **norms of Riesz and Bessel potentials of orthonormal functions** [7]. Again, $\phi_1, \dots, \phi_N$ are L^2 orthonormal and let

$$u_i = (-\Delta + m^2)^{-1/2} \phi_i \quad (22)$$

$$\rho(x) \equiv \sum_{i=1}^N |u_i(x)|^2. \quad (23)$$

Then there are constants L, B_p, A_n (independent of m) such that

$$n = 1 : \|\rho\|_\infty \leq L/m, \quad m > 0 \quad (24)$$

$$n = 2 : \|\rho\|_p \leq B_p m^{-2/p} N^{1/p}, \quad 1 \leq p < \infty, m > 0 \quad (25)$$

$$n \geq 3 : \|\rho\|_p \leq A_n N^{1/p}, \quad p = n/(n-2), m \geq 0. \quad (26)$$

If the orthogonality condition is dropped then the right sides of (24)-(26) have to be multiplied by $N, N^{1-1/p}, N^{1-1/p}$ respectively. Possibly the absence of N in (24) is the most striking. Similar results can be derived [7] for $(-\Delta + m^2)^{-\alpha/2}$ in place of $(-\Delta + m^2)^{-1/2}$, with $\alpha < n$ when $m = 0$.

Inequality (15) also has applications in mathematical physics.

Application 2. (Navier-Stokes equation.) Suppose $\Omega \subset \mathbf{R}^n$ is an open set with finite volume $|\Omega|$ and consider

$$H = -\Delta - V(x)$$

on Ω with Dirichlet boundary conditions. Let $\lambda_1 \leq \lambda_2 \leq \dots$ be the eigenvalues of H . Let $\tilde{N}$ be the smallest integer, N , such that

$$E_N \equiv \sum_{i=1}^N \lambda_i \geq 0. \quad (27)$$

We want to find an upper bound for $\tilde{N}$.

If $\phi_1, \phi_2, \dots$ are the normalized eigenfunctions then, from (13)-(15) with $\phi_1, \dots, \phi_N$,

$$E_N = T - \int \rho V \geq K_n \int \rho^{1+n/2} - \int V_+ \rho \geq G(\rho), \quad (28)$$

where (with $p = 1 + n/2$ and $q = 1 + 2/n$)

$$G(\rho) \equiv K_n \|\rho\|_p^p - \|V_+\|_q \|\rho\|_p. \quad (29)$$

Thus, for all N ,

$$E_N \geq \inf\{G(\rho) \mid \|\rho\|_1 = N, \rho(x) \geq 0\}. \quad (30)$$

But $\|\rho\|_p |\Omega|^{1/q} \geq \|\rho\|_1 = N$ so, with $X \equiv \|\rho\|_p$,

$$E_N \geq \inf\{J(X) \mid X \geq N|\Omega|^{-1/q}\} \quad (31)$$

where

$$J(X) \equiv K_n X^p - \|V_+\|_q X. \quad (32)$$

Now $J(X) \geq 0$ for $X \geq X_0 = \{\|V_+\|_q/K_n\}^{1/(p-1)}$, whence we have the following implication:

$$N \geq |\Omega|^{1/q} \{\|V_+\|_q/K_n\}^{1/(p-1)} \implies E_N \geq 0. \quad (33)$$

Therefore

$$\bar{N} \leq |\Omega|^{1/q} \{\|V_+\|_q/K_n\}^{1/(p-1)}. \quad (34)$$

The bound (34) can be applied [8] (following an idea of Ruelle) to the Navier-stokes equation. There, $\bar{N}$ is interpreted as the Hausdorff dimension of an attracting set for the N-S equation, while $V(x) \equiv \nu^{-3/2} \varepsilon(x)$, where $\varepsilon(x) = \nu |\nabla v(x)|^2$ is the average energy dissipation per unit mass in a flow v . ν is the viscosity.

Application 3. (Stability of matter.) This is the original application [10,11]. In the quantum mechanics of Coulomb systems (electrons and nuclei) one wants a lower bound for the Hamiltonian operator:

$$\begin{aligned} H = & - \sum_{i=1}^N \Delta_i - \sum_{i=1}^N \sum_{j=1}^K z_j |x_i - R_j|^{-1} + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} \\ & + \sum_{1 \leq i < j \leq K} z_i z_j |R_i - R_j|^{-1} \end{aligned} \quad (35)$$

on the L^2 space of *antisymmetric* functions $\psi(x_1, \dots, x_N), x_i \in \mathbf{R}^3$. Here, N is the number of electrons (with coordinates x_i) and $R_1, \dots, R_K \in \mathbf{R}^3$ are fixed vectors

representing the locations of fixed nuclei of charges $z_1, \dots, z_K > 0$. The desired bound is linear:

$$H \geq -A(N + K) \quad (36)$$

for some A independent of $N, K, R_1, \dots, R_K$ (assuming all $z_i < \text{some } \bar{z}$).

The main point is that antisymmetry of ψ is crucial for (36) and this is reflected in the fact that (15) holds with antisymmetry, but only (17) holds without it. Without the antisymmetry condition, H would grow as $-(N + K)^{5/3}$. This is discussed in Application 6 below. By using (15) one can eliminate the differential operators Δ_i . The functional $\psi \rightarrow (\psi, H\psi)$, with $(\psi, \psi) = 1$ can be bounded below using (15) by a functional (called the Thomas-Fermi functional) involving only $\rho(x)$ defined in (21). The minimization of this latter functional with respect to ρ is tractable and leads to (36).

Application 4. (Stellar structure.) Going from atoms to stars, we now consider N neutrons which attract each other gravitationally with a coupling constant $\kappa = Gm^2$, where G is the gravitational constant and m is the neutron mass. There are no Coulomb forces. Moreover, a “relativistic” form is assumed for the kinetic energy, which means that $-\Delta$ is replaced by $(-\Delta)^{1/2}$. Thus (35) is replaced by

$$H_N = \sum_{i=1}^N (-\Delta_i)^{1/2} - \kappa \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} \quad (37)$$

(again on antisymmetric functions). One finds asymptotically for large N , that

$$\begin{aligned} \inf \text{spec}(H_N) &= 0 \quad \text{if } \kappa \leq CN^{-2/3} \\ &= -\infty \quad \text{if } \kappa > CN^{-2/3} \end{aligned} \quad (38)$$

for some constant, C . Without antisymmetry, $N^{-2/3}$ must be replaced by N^{-1} . Equation (38) is proved in [12]. An important role is played by Daubechies’s generalization [4] of (15) to the operator $(-\Delta)^{1/2}$ on $L^2(\mathbf{R}^N)$, namely (for antisymmetric ψ with $\|\psi\|_2 = 1$)

$$\left(\psi, \sum_{i=1}^N (-\Delta_i)^{1/2} \psi \right) \geq B_n \int \rho(x)^{1+1/n} \quad (39)$$

with ρ given by (19), (21). In general, one has

$$\left(\psi, \sum_{i=1}^N (-\Delta)^p \psi \right) \geq C_{p,n} \int \rho(x)^{1+2p/n} dx. \quad (40)$$

Recently [13] there has been considerable progress in this problem beyond that in [12]. Among other results there is an evaluation of the sharp asymptotic C in (38), i.e. if we first define $\kappa^c(N)$ to be the precise value of κ at which $\inf \text{spec}(H_N) = -\infty$, we then define

$$C = \lim_{N \rightarrow \infty} N^{2/3} \kappa^c(N). \quad (41)$$

Let B_n^c be the "classical guess" in (39). This can be calculated from the analogue of (9) (using $|p|$ instead of p^2 , and which leads to $\sum |e_i| \approx \tilde{L} \int V_+^{1+n}$), and then from the analogue of (16), namely $\tilde{L} = C_n B_n^{-n}$. One finds $B_3^c = (3/4)(6\pi^2)^{1/3}$ (cf. (18)). Using B_3^c , we introduce the functional

$$\mathcal{E}(\rho) = B_3^c \int \rho^{4/3} - \frac{1}{2} \kappa \int \int \rho(x) \rho(y) |x - y|^{-1} dx dy \quad (42)$$

for $\rho \in L^1(\mathbf{R}^3) \cap L^{4/3}(\mathbf{R}^3)$ and define the energy

$$E^c(N) = \inf \{ \mathcal{E}(\rho) \mid \int \rho = N \}. \quad (43)$$

One finds there is a finite $\alpha^c > 0$ such that $E^c(N) = 0$ if $\kappa N^{2/3} \leq \alpha^c$ and $E^c(N) > -\infty$ if $\kappa N^{2/3} > \alpha^c$. (This α^c is found by solving a Lane-Emden equation.)

Now (42) and (43) constitute the semiclassical approximation to H_N in the following sense. We expect that if we set $\kappa = \alpha N^{-2/3}$ in (37), with α fixed, then if $\alpha < \alpha^c$

$$\lim_{N \rightarrow \infty} \inf \text{spec}(H_N) = 0 \quad (44)$$

while if $\alpha > \alpha^c$ there is an N_0 such that

$$\inf \text{spec}(H_N) = -\infty \quad \text{if } N > N_0. \quad (45)$$

Indeed, (44) and (45) are true [13], and thus α^c is the sharp asymptotic value of C in (38).

An interesting point to note is that Daubechies's B_3 in (39) is about half of B_3^c . The sharp value of B_3 is unknown. Nevertheless, with some additional tricks one can get from (37) to (42) with B_3^c and not B_3 . Inequality (39) plays a role in [13], but it is not sufficient.

Application 5. (Stability of atoms in magnetic fields.) This is given in [9]. Here $\psi(x_1, \dots, x_N)$ becomes a spinor-valued function, i.e. ψ is an antisymmetric function in $\bigwedge_1^N L^2(\mathbf{R}^3; \mathbf{C}^2)$. The operator H of interest is as in (35) but with the replacement

$$-\Delta \rightarrow \{ \sigma \cdot (i \nabla - A(x)) \}^2 \quad (46)$$

where $\sigma_1, \sigma_2, \sigma_3$ are the 2×2 Pauli matrices (i.e. generators of $SU(2)$) and $A(x)$ is a given vector field (called the magnetic vector potential). Let

$$E_0(A) = \inf \text{spec}(H) \quad (47)$$

after the replacement of (46) in (35). As $A \rightarrow \infty$ (in a suitable sense), $E_0(A)$ can go to $-\infty$. The problem is this: Is

$$\tilde{E}(A) \equiv E_0(A) + \frac{1}{8\pi} \int (\text{curl } A)^2 \quad (48)$$

bounded below for all A ? In [9] the problem is resolved for $K = 1$, all N and $N = 1$, all K . It turns out that $\tilde{E}(A)$ is bounded below in these cases if and only if all the z_i satisfy $z_i < z^c$ where z^c is some fixed constant independent of N and K . The problem is still open for all N and all K .

One of the main problems in bounding $\tilde{E}(A)$ is to find a lower bound for the kinetic energy (the first term in (35) after the replacement given in (46)) for an antisymmetric ψ . First, there is the identity

$$\left(\psi, \sum_{i=1}^N \{ \sigma \cdot (i\nabla - A(x_i)) \}^2 \psi \right) = T(\psi, A) - \left(\psi, \sum_{i=1}^N \sigma \cdot B(x_i) \psi \right) \quad (49)$$

with $B = \text{curl } A$ being the magnetic field and

$$T(\psi, A) = \left(\psi, \sum_{i=1}^N |i\nabla - A(x)|^2 \psi \right). \quad (50)$$

The last term on the right side of (49) can be controlled, so it will be ignored here. The important term is $T(\psi, A)$. Since Pauli matrices do not appear in (50) we can now let ψ be an ordinary complex valued (instead of spinor valued) function.

It turns out that (8), and hence (15), hold with some $\tilde{L}_{\gamma,n}$ which is *independent* of A . The T in (15) is replaced, of course, by the $T(\psi, A)$ of (50). To be more precise, the *sharp* constants $L_{\gamma,n}$ and $\tilde{L}_{\gamma,n}$ are unknown (except for $\gamma \geq 3/2, n = 1$ in the case of $L_{\gamma,n}$) and conceivably $\tilde{L}_{\gamma,n} > L_{\gamma,n}$. However, all the current bounds for $L_{\gamma,n}$ (see [8]) also hold for $\tilde{L}_{\gamma,n}$. Thus, for $n = 3$ we have

$$T(\psi, A) \geq K_3 \int \rho^{5/3} \quad (51)$$

with K_3 being the value given in [8], namely 2.7709.

However, in [9] another inequality is needed

$$T(\psi, A) \geq C \left\{ \int \rho^2 \right\}^{2/3}. \quad (52)$$

It seems surprising that we can go from an $L^{5/3}$ estimate to an L^2 estimate, but the surprise is diminished if (17a) with its L^3 estimate is recalled. First note that (1) holds (with the same S_n) if $|\nabla f|^2$ is replaced by $||[i\nabla - A(x)]f|^2$. (By writing $f = |f|e^{i\theta}$ one finds that $|\nabla|f||^2 \leq |[i\nabla - A(x)]f|^2$.) Then (17a) holds since only convexity was used. Thus, using the mean of (15) and (17a),

$$T(\psi, A) \geq (S_n K_n)^{1/2} \|\rho\|_3^{1/2} \|\rho\|_{5/3}^{5/6}. \quad (53)$$

An application of Hölders inequality yields (52) with $C^2 = S_n K_n$.

Application 6. (Instability of bosonic matter.) As remarked in Application 3, dropping the antisymmetry requirement on ψ (the particles are now bosons) makes $\inf \text{spec}(H)$ diverge as $-(N + K)^{5/3}$. The extra power $2/3$, relative to (36) can be traced directly to the factor $N^{-2/3}$ in (17).

An interesting problem is to allow the positive particles also to be movable and to have charge $z_i = 1$. This should raise $\inf \text{spec} H$, but by how much? For $2N$ particles the new H is

$$\tilde{H} = - \sum_{i=1}^{2N} \Delta_i + \sum_{1 \leq i < j \leq 2N} e_i e_j |x_i - x_j|^{-1} \quad (54)$$

with $e_i = +1$ for $1 \leq i \leq N$ and $e_i = -1$ for $N + 1 \leq i \leq 2N$. $\tilde{H}$ acts on $L^2(\mathbf{R}^{3N})$ without any symmetry requirements.

Twenty years ago Dyson [5] proved, by a variational calculation, that $\inf \text{spec}(\tilde{H}) \leq -AN^{7/5}$ for some $A > 0$. Thus, stability (i.e. a linear law (36)) is not restored, but the question of whether the correct exponent is $7/5$ or $5/3$, or something in between, remained open. It has now been proved [2] that $N^{7/5}$ is correct, $\inf \text{spec}(\tilde{H}) \geq -BN^{7/5}$. The proof is much harder than for (36) because no simple semiclassical theory (like Thomas-Fermi theory) is a good approximation to $\tilde{H}$. Correlations are crucial.

Application 7. (Stability of relativistic matter.) Let us return to Coulomb systems (electrons and nuclei) as in application 3, but with (35) replaced by

$$H = \sum_{i=1}^N \{(-\Delta_i + m^2)^{1/2} - m\} + \alpha V_c(x_1, \dots, x_N; R_1, \dots, R_K) \quad (55)$$

with $\alpha = e^2$ = electron charge squared (and $\hbar = c = 1$) and where

$$V_c = - \sum_{i=1}^N \sum_{j=1}^K z_j |x_i - R_j|^{-1} + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} + \sum_{1 \leq i < j \leq K} z_i z_j |R_i - R_j|^{-1} \quad (56)$$

is the Coulomb potential. The electron charge, $\alpha^{1/2}$, is explicitly displayed in (55) for a reason to be discussed presently. Also (55) differs from (35) in that the kinetic energy operator $-\Delta$ is replaced by the relativistic form $(-\Delta + m^2)^{1/2} - m$, where m is the electron mass. Since $-\Delta - m \leq (-\Delta + m^2)^{1/2} - m \leq -\Delta$, the difference of these two operators is a bounded operator and therefore, as far as the stability question is concerned, we may as well use the simplest operator $(-\Delta)^{1/2}$ in (55), which will be done henceforth. This, in fact, was already done in (37). We define

$$E_{N,K}(R_1, \dots, R_K) = \inf \text{spec} H \quad (57)$$

$$E_{N,K} = \inf_{R_1, \dots, R_K} E_{NK}(R_1, \dots, R_K) \quad (58)$$

$$E = \inf_{N,K} E_{NK} \quad (59)$$

Under scaling (dilation of coordinates in $\mathbf{R}^{3N+3K}$) the operators $(-\Delta)^{1/2}$ and $|x|^{-1}$ behave the same (proportional to length) $^{-1}$ and hence we conclude that

$$E_{N,K} = 0 \quad \text{or} \quad -\infty. \quad (60)$$

The system is said to be **stable** if $E = 0$. For simplicity of exposition let us take all z_j to be some common value, z .

For the hydrogenic atom $N = K = 1$ the only constant that appears is the combination $z\alpha$. It is known that $E_{1,1} = 0$ if and only if $z\alpha \leq 2/\pi$. In the many-body case there are two constants (which can be taken to be $z\alpha$ and α) and the question is whether the system is stable *all the way up to* $z\alpha = 2/\pi$ for α less than some small, but fixed $\alpha_c > 0$. The answer will depend on q , the number of spin states allowed for the fermionic electrons. (Note: in application 3 we implicitly took $q = 1$. In fact $q = 2$ in nature. To say that there are q spin states means that under permutations $\psi(x_1, \dots, x_N)$ belongs to a Young's tableaux of q or fewer columns.)

This problem is resolved in [15] where it is shown that stability occurs if

$$q\alpha \leq 1/47. \quad (61)$$

The kinetic energy bound (39) plays a crucial role in the proof (but, of course, many other inequalities are also needed). It is also shown in [15] that stability definitely fails to occur if

$$\alpha > 36q^{-1/3} z^{2/3} \quad (62)$$

or if

$$\alpha > 128/15\pi. \quad (63)$$

If (63) holds then instability occurs for *every* $z > 0$, no matter how small.

REFERENCES

- [1] M. Aizenman and E.H. Lieb, On semiclassical bounds for eigenvalues of Schrödinger operators, *Phys. Lett.* 66A, 427-429 (1978).
- [2] J. Conlon, E.H. Lieb and H.-T. Yau, The $N^{7/5}$ law for bosons, *Commun. Math. Phys.* (submitted).
- [3] M. Cwikel, Weak type estimates for singular values and the number of bound states of Schrödinger operators, *Ann. Math.* 106, 93-100 (1977).
- [4] I. Daubechies, *Commun. Math. Phys.* 90, 511-520 (1983).
- [5] F.J. Dyson, Ground state energy of a finite systems of charged particles, *J. Math. Phys.* 8, 1538-1545 (1967).
- [6] E.H. Lieb, The number of bound states of one-body Schrödinger operators and the Weyl problem, *A.M.S. Proc. Symp. in Pure Math.* 36, 241-251 (1980). The results were announced in *Bull. Ann. Math. Soc.* 82, 751-753 (1976).
- [7] E.H. Lieb, An L^p bound for the Riesz and Bessel potentials of orthonormal functions, *J. Funct. Anal.* 51, 159-165 (1983).
- [8] E.H. Lieb, On characteristic exponents in turbulence, *Commun. Math. Phys.* 92, 473-480 (1984).
- [9] E.H. Lieb and M. Loss, Stability of Coulomb systems with magnetic fields: II. The many-electron atom and the one-electron molecule, *Commun. Math. Phys.* 104, 271-282 (1986).
- [10] E.H. Lieb and W.E. Thirring, Bounds for the kinetic energy of fermions which proves the stability of matter, *Phys. Rev. Lett.* 35, 687-689 (1975). Errata 35, 1116 (1975).
- [11] E.H. Lieb and W.E. Thirring, "Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities" in *Studies in Mathematical Physics* (E. Lieb, B. Simon, A. Wightman eds.) Princeton University Press, 1976, pp. 269-304.
- [12] E.H. Lieb and W.E. Thirring, Gravitational collapse in quantum mechanics with relativistic kinetic energy, *Ann. of Phys. (NY)* 155, 494-512 (1984).
- [13] E.H. Lieb and H.-T. Yau, The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics, *Commun. Math. Phys.* 112, 147-174 (1987).
- [14] G.V. Rosenbljum, Distribution of the discrete spectrum of singular differential operators. *Dokl. Akad. Nauk SSSR* 202, 1012-1015 (1972). (MR 45 #4216). The details are given in *Izv. Vyss. Ucebn. Zaved. Matem.* 164, 75-86 (1976). (English trans. *Sov. Math. (Iz VUZ)* 20, 63-71 (1976).)
- [15] E.H. Lieb and H.T. Yau, The stability and instability of relativistic matter, *Commun. Math. Phys.* 118, 177-213 (1988). For a short summary see: Many body stability implies a bound on the fine structure constant, *Phys. Rev. Lett.* 61, 1695-1697 (1988).

INEQUALITIES FOR THE MOMENTS OF THE EIGENVALUES OF THE SCHRÖDINGER HAMILTONIAN AND THEIR RELATION TO SOBOLEV INEQUALITIES

Elliott H. Lieb*
Walter E. Thirring

1. Introduction

Estimates for the number of bound states and their energies, $e_j \leq 0$, are of obvious importance for the investigation of quantum mechanical Hamiltonians. If the latter are of the single particle form $H = -\Delta + V(x)$ in $\mathbb{R}^n$, we shall use available methods to derive the bounds

$$\sum_j |e_j|^\gamma \leq L_{\gamma,n} \int d^n x |V(x)|_-^{\gamma+n/2}, \quad \gamma > \max(0, 1-n/2). \quad (1.1)$$

Here, $|V(x)|_- = -V(x)$ if $V(x) \leq 0$ and is zero otherwise.

Of course, in many-body theory, one is more interested in Hamiltonians of the form $-\sum_i \Delta_i + \sum_{i>j} v(x_i - x_j)$. It turns out, however, that the energy bounds for the single particle Hamiltonian yield a lower bound for the kinetic energy, T , of N fermions in terms of integrals over the single particle density defined by

$$\rho(x) \equiv N \int |\psi(x, x_2, \dots, x_N)|^2 d^n x_2 \dots d^n x_N, \quad (1.2)$$

where ψ is an antisymmetric, normalized function of the N variables $x_i \in \mathbb{R}^n$. Our main results, in addition to (1.1), will be of the form

* Work supported by U. S. National Science Foundation Grant MPS 71-03375-A03.

$$\begin{aligned}
 T &\equiv \sum_{i=1}^N \int |\vec{\nabla}_i \psi(x_1 \cdots x_N)|^2 d^n x_1 \cdots d^n x_N \\
 &\geq K_{p,n} \left[\int d^n x \rho(x)^{p/(p-1)} \right]^{2(p-1)/n}
 \end{aligned} \tag{1.3}$$

when $\max\{n/2, 1\} \leq p \leq 1 + n/2$.

For $N = 1$, $p = n/2$, (1.3) reduces to the well-known Sobolev inequalities. (1.3) is therefore a partial generalization of these inequalities, and we shall expand on this in Section 3.

Our constants $K_{p,n}$ are not always the best possible ones, but nevertheless, they may be useful for many purposes. In particular, in ref. [1], a special case of (1.3) was used to give a simple proof of the stability of matter, with a constant of the right order of magnitude. The result for q species of fermions ($2m = e = \hbar = 1$) moving in the field of M nuclei with positive charges Z_j is

$$H \geq -1.31 q^{2/3} N \left[1 + \left(\sum_{j=1}^M Z_j^{7/3} / N \right)^{1/2} \right]^2. \tag{1.4}$$

In particular, if $q = 2$ (spin 1/2 electrons), we have a bound $\sim N$, and if we set $q = N$, we get a bound $\sim N^{5/3}$ if no symmetry requirement is imposed on the wave function; *a fortiori* this is a bound for *bosons*. Our bound implies stability of matter in its intuitive meaning such that the volume occupied by N particles will be $\sim N$ (Bohr radius)³. To give a formal demonstration of this fact, one might use a method which gives lower bounds for the radii of complex atoms (compare Equation (3.6, 38) of ref. [20]). As a first observation, one calculates the ground state energy of N electrons (with spin) in a harmonic potential. Filling the oscillator levels, one finds

$$\frac{1}{2} \sum_{i=1}^N (-\Delta_i + \omega^2 \vec{x}_i^2) \geq \omega N^{4/3} \frac{3^{4/3}}{4} (1 + O(N^{-1/3})). \tag{1.5}$$

Next, take the expectation value of this operator inequality with the ground state of H , set

$$\omega = \frac{4}{3^{4/3}} < - \sum_i \Delta_i > \frac{1}{N^{4/3}} \quad (1.6)$$

and use the virial theorem

$$< - \sum_i \Delta_i > = -E_0 \leq 2.08 N \left[1 + \left(\sum_{j=1}^M Z_j^{7/3}/N \right)^{1/2} \right]^2. \quad (1.7)$$

Altogether we find

$$< \sum_{i=1}^N \vec{x}_i^2 > \geq \frac{(3N)^{8/3}}{16 < - \sum_i \Delta_i >} \geq \frac{3^{8/3} N^{5/3}}{16 \cdot 2.08 \left[1 + \left(\sum_j Z_j^{7/3}/N \right)^{1/2} \right]^2}. \quad (1.8)$$

Thus we have proved that

$$< \vec{x}_i^2 >^{1/2} \geq c N^{1/3}, \quad c = \frac{.75}{\left[1 + \left(\sum_j Z_j^{7/3}/N \right)^{1/2} \right]}. \quad (1.9)$$

Therefore, if the system is not compressed by other forces, so that the virial theorem is valid, it will not collapse, but will adjust its volume to a size proportional to the number of particles. Regarding the Z -dependence, we see that with $Z = Z_j = N/M$ we have (for large Z)

$$< \vec{x}_i^2 >^{1/2} \sim M^{1/3} Z^{-1/3}.$$

That is, the mean atomic radius is predicted to be $\geq Z^{-1/3}$. A better result can hardly be expected since for $M = 1$, this is the correct Z -dependence for large Z .

Although we have no results on the best possible constants, $K_{p,n}$, except in a few special cases, experience drawn from computer calculations suggests that there is a critical value $\gamma_{c,n}$ above which the classical value gives a bound:

$$\begin{aligned}
 \left(\sum_j |e_j|^\gamma \right)_{\text{classical}} &= (2\pi)^{-n} \int d^n p \, d^n x \, |p^2 + V(x)|_-^\gamma \\
 &\equiv L_{\gamma,n}^C \int |V(x)|_-^{\gamma+n/2} d^n x \, , \\
 \gamma &\geq \gamma_{C,n} \, ,
 \end{aligned} \tag{1.10}$$

and where $L_{\gamma,n}^C$, given by the above integral, is

$$L_{\gamma,n}^C = 2^{-n} \pi^{-n/2} \Gamma(\gamma+1)/\Gamma(\gamma+1+n/2) \, . \tag{1.11}$$

We conjecture $\gamma_{C,1} = 3/2$, $\gamma_{C,3} \cong .863$ and $\gamma_{C,n} = 0$, all $n \geq 8$. If this conjecture were to be true, the constants in (1.3, 1.4) could be further improved.

In the next section we shall deduce bounds for $\sum_j |e_j|^\gamma$ and use them in Section 3 to derive (1.3). In Section 4 we shall discuss our conjectures and support them for $n = 1$ with results from the Korteweg-de Vries equation. Section 5 contains new results added in proof. In Appendix A, generously contributed by J. F. Barnes, further evidence from computer studies is presented. We are extremely grateful to Dr. Barnes for taking an interest in this problem, for without his results we would have been hesitant to put forth our conjectures.

2. Bounds for Moments of the Eigenvalues

In this section we shall deduce bounds of the form (1.1), and we shall compare our $L_{\gamma,n}$ with the classical values which one gets by replacing

$$\sum_j |e_j|^\gamma \quad \text{by} \quad (2\pi)^{-n} \int d^n x \, d^n p \, |p^2 + V(x)|_-^\gamma \, .$$

For $n \sim 3$ and $\gamma \sim 1$, the latter are smaller by about an order of magnitude.

Our inequalities are based on the Birman-Schwinger [2, 3] method for estimating N_E , the number of bound states of $H = -\Delta + V(x)$ having an energy $\leq E$. Since

$$\frac{\partial}{\partial E} N_E = \sum_j \delta(E - e_j)$$

we have

$$\sum_j |e_j|^\gamma = \gamma \int_0^\infty da a^{\gamma-1} N_{-a} . \quad (2.1)$$

Now, according to Birman-Schwinger [2, 3], for all $a \geq 0$, $m \geq 1$ and $t \in [0, 1]$,

$$N_{-a} \leq \text{Tr}(|V + (1-t)a|_-^{1/2} (-\Delta + ta)^{-1} |V + (1-t)a|_-^{1/2})^m . \quad (2.2)$$

REMARKS ABOUT (2.2):

1. We are only interested in potentials such that $V_- \in L^{\gamma+n/2}(\mathbb{R}^n)$ for $\gamma \geq \min(0, 1 - n/2)$. For such potentials (2.2) is justified, and a complete discussion is given in Simon [4, 5]. Moreover, it is sufficient to consider $V \in C_0^\infty(\mathbb{R}^n)$ in (2.2), and in the rest of this paper, and then to use a limiting argument. Such potentials have the advantage that they have only a finite number of bound states [5].

2. Since we are interested in maximizing $\sum |e_j|^\gamma / \int |V|_-^{\gamma+n/2}$, we may as well assume that $V(x) \leq 0$, i.e. $V = -|V|_-$. This follows from the max-min principle [4] which asserts that $e_j(V) \geq e_j(-|V|_-)$, all j , including multiplicity.

To evaluate the trace in (2.2), we use the inequality

$$\text{Tr}(B^{1/2} A B^{1/2})^m \leq \text{Tr} B^{m/2} A^m B^{m/2} \quad (2.3)$$

when A, B are positive operators and $m \geq 1$. When m is integral and A, B is of our special form, (2.3) is a consequence of Hölder's inequality. For completeness, we shall give a more general derivation of (2.3) in Appendix B.

To calculate

$$\text{Tr} |V + (1-t)\alpha|_-^m (-\Delta + t\alpha)^{-m}, \quad (2.4)$$

we shall use an x -representation where $(-\Delta + t\alpha)^{-m}$ is the kernel

$$G_{t\alpha}^{(m)}(x-y) = (2\pi)^{-n} \int d^n p (p^2 + t\alpha)^{-m} e^{ip(x-y)} \quad (2.5)$$

if $m > n/2$. Using

$$d^n p = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty dp p^{n-1}, \quad (2.6)$$

we easily compute

$$\begin{aligned} G_{t\alpha}^{(m)}(0) &= (2\pi)^{-n} \frac{2\pi^{n/2}}{\Gamma(n/2)} (t\alpha)^{-m+n/2} \int_0^\infty p^{n-1} (p^2 + 1)^{-m} dp \\ &= (4\pi)^{-n/2} \frac{\Gamma(m-n/2)}{\Gamma(m)} (t\alpha)^{-m+n/2} \end{aligned} \quad (2.7)$$

if $m > n/2$. Thus,

$$N_{-\alpha} \leq (4\pi)^{-n/2} \frac{\Gamma(m-n/2)}{\Gamma(m)} (t\alpha)^{-m+n/2} \int d^n x |V(x) + (1-t)\alpha|_-^m. \quad (2.8)$$

Next, we substitute (2.8) into (2.1). If we impose the condition that $t < 1$, it is easy to prove that one can interchange the α and the x integration. Changing variables $\alpha \rightarrow (1-t)^{-1} |V(x)|_- \beta$, leads to

$$\begin{aligned} \sum |e_j|^\gamma &\leq \gamma (4\pi)^{-n/2} t^{-m+n/2} (1-t)^{m-\gamma-n/2} \frac{\Gamma(\gamma-m+n/2) \Gamma(m-n/2)}{\Gamma(\gamma+1+n/2)} m \int d^n x \\ &\quad \times |V(x)|_-^{\gamma+n/2} \end{aligned} \quad (2.9)$$

provided $n/2 < m < n/2 + \gamma$, $m \geq 1$ and $0 < t < 1$. The optimal t is $t = (m-n/2)/\gamma$.

If we put our results together, we obtain the following (see note added in proof, Section 5).

THEOREM 1. *Let $V_- \in L^{\gamma+n/2}(\mathbb{R}^n)$, $\gamma \geq \max(0, 1-n/2)$. Let $H = -\Delta + V(x)$, and let $e_j \leq 0$ be the negative energy bound states of H . Then*

$$\sum |e_j|^\gamma \leq L_{\gamma,n} \int |V(x)|_-^{\gamma+n/2} \quad (2.10)$$

where

$$L_{\gamma,n} \leq \tilde{L}_{\gamma,n} \equiv \min_m (4\pi)^{-n/2} \gamma^{\gamma+1} \frac{m}{\Gamma(\gamma + \frac{n}{2} + 1)} F\left(m - \frac{n}{2}\right) F\left(\gamma + \frac{n}{2} - m\right), \quad (2.11)$$

and where $F(x) = \Gamma(x) x^{-x}$, $\max\{1, n/2\} \leq m < n/2 + \gamma$.

REMARKS:

1. When $\gamma = 0$, $\sum |e_j|^0$ means the number of bound states, including zero energy states. For $n \geq 2$, our $\tilde{L}_{0,n} = \infty$. In Section 4, we shall discuss the $\gamma = 0$ case further. See also Section 5.

2. In (2.11), $\tilde{L}_{\gamma,n}$ is the bound we have obtained using the Birman-Schwinger principle. We shall henceforth reserve the symbol $L_{\gamma,n}$ for the quantity

$$L_{\gamma,n} \equiv \sup_V \sum |e_j|^\gamma / \int |V|_-^{\gamma+n/2}. \quad (2.12)$$

Optimization with respect to m in (2.11) can be done either numerically or analytically in the region where Stirling's formula

$$F(x) \sim e^{-x} \sqrt{2\pi/x} \quad (2.13)$$

can be applied. In [1], for $n = 3$, $\gamma = 1$, we used the value 2 for m . A marginal improvement can be obtained with $m = 1.9$.

If (2.13) were exact, the best m would be

$$\bar{m} = n(\gamma + n/2)/(n + \gamma). \quad (2.14)$$

Note that as $\gamma \rightarrow \infty$, $\bar{m}$ is bounded by n . Using $\bar{m}$, together with (2.13), which is valid when $\gamma n(\gamma + n)^{-1}$ is large,

$$\tilde{L}_{\gamma,n} \sim (4\pi)^{1-n/2} \frac{\gamma^\gamma e^{-\gamma}}{\Gamma(\gamma+n/2)} \left[\frac{n/2}{\gamma+n/2} \right]^{1/2}. \quad (2.15)$$

Finally, we want to compare our bounds with their classical values, $L_{\gamma,n}^C$. From the results of Martin [6] and Tamura [7], one has the following

THEOREM 2. *If $V(x) \leq 0$ and $V \in C_0^\infty(\mathbb{R}^n)$, then*

$$\lim_{\lambda \rightarrow \infty} \sum_j |e_j(\lambda V)|^\gamma / \int |\lambda V|^{\gamma+n/2} = L_{\gamma,n}^C. \quad (2.16)$$

COROLLARY.

$$L_{\gamma,n} \geq L_{\gamma,n}^C. \quad (2.17)$$

Our $\tilde{L}_{\gamma,n}$ satisfies (2.17), in particular in the asymptotic region (2.15), we find

$$\tilde{L}_{\gamma,n}/L_{\gamma,n}^C \approx [4\pi n(\gamma+n/2)]^{1/2} \gamma^{-1/2}. \quad (2.18)$$

We conjecture in Section 4 that for γ sufficiently large, the best possible $L_{\gamma,n}$ should be $L_{\gamma,n}^C$, a result which does not follow from the Birman-Schwinger method employed here. For small γ , we know that $L_{\gamma,n}^C$ is not a bound.

We conclude this section with a theorem about $L_{\gamma,n}$ which will be useful in the discussion of the one-dimensional case in Section 4.

THEOREM 3. *Let $\gamma \geq 1 + \max(0, 1 - n/2)$. Then*

$$L_{\gamma,n} \leq L_{\gamma-1,n} [\gamma/(\gamma+n/2)]. \quad (2.19)$$

PROOF. Choose $\varepsilon > 0$. We can find a $V \in C_0^\infty(\mathbb{R}^n)$, with $V \leq 0$, such that

$$L_{\gamma,n}(V) = \sum_j |e_j(V)|^\gamma / \int |V|^{\gamma+n/2} \geq L_{\gamma,n} - \varepsilon.$$

Let $g \in C_0^\infty(\mathbb{R}^n)$ be such that $0 \leq g(x) \leq 1$, $\forall x$, and $V(x) \neq 0$ implies $g(x) = 1$. Let $V_\lambda(x) = V(x) - \lambda g(x)$, $\lambda \leq 0$. The functions $|e_j(V_\lambda)|$ are continuous and monotone increasing in λ . Furthermore, there are a finite number of values $-\infty < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k \leq 0$ with λ_j being the value of λ at which $e_j(V_\lambda)$ first appears. λ_1 is finite because V_λ is non-negative for λ sufficiently negative. $e_j(V_\lambda)$ is continuously differentiable on $\Lambda = \{\lambda | 0 \geq \lambda > \lambda_1, \lambda \neq \lambda_i, i = 1, \dots, k\}$ and

$$de_j(V_\lambda)/d\lambda = - \int |\psi_j(x; V_\lambda)|^2 g(x) d^n x$$

by the Feynman-Hellman theorem. It is easy to prove that if $f, g \in L^p(\mathbb{R}^n)$, $p > 1$, then

$$h(\lambda) \equiv \int |f(x) - \lambda g(x)|_-^p d^n x$$

is differentiable, $\forall \lambda$ and

$$dh/d\lambda|_{\lambda=0} = p \int |f(x)|_-^{p-1} g(x) d^n x.$$

Thus $L_{\gamma,n}(V_\lambda)$ is piecewise C^1 on Λ and its derivative, $\dot{L}_{\gamma,n}$, is given by

$$\dot{L}_{\gamma,n} = \left[\int |V_\lambda|_-^{\gamma+n/2} \right]^{-1} \left\{ \gamma \sum_j |e_j(V_\lambda)|^{\gamma-1} \int g(x) |\psi_j(x; V_\lambda)|^2 d^n x - (\gamma+n/2) L_{\gamma,n}(V_\lambda) \right. \\ \left. \cdot \int |V_\lambda(x)|_-^{\gamma+n/2-1} g(x) d^n x \right\}.$$

By the stated properties of $L_{\gamma,n}$, there exists a $\lambda \in (\lambda_1, 0]$ such that

- (i) $\dot{L}_{\gamma,n}(V_\lambda) \geq 0$;
- (ii) $L_{\gamma,n}(V_\lambda) \geq L_{\gamma,n} - 2\epsilon$.

Thus, using the properties of g ,

$$0 \leq \gamma \sum_j |e_j(V_\lambda)|^{\gamma-1} - L_{\gamma,n}(V_\lambda)(\gamma+n/2) \int |V_\lambda|_-^{\gamma+n/2-1}. \quad (2.20)$$

Since ϵ was arbitrary, (2.20) implies the theorem.

If we use (2.17) together with the fact that $L_{\gamma,n}^C = L_{\gamma-1,n}^C[\gamma/(\gamma+n/2)]$, we have

COROLLARY. *If for some $\gamma \geq \max(0, 1-n/2)$, $L_{\gamma,n} = L_{\gamma,n}^C$, then*

$$L_{\gamma+j,n} = L_{\gamma+j,n}^C, \quad j = 0, 1, 2, 3, \dots$$

REMARK. By the same proof

$$L_{\gamma,n}^1 \leq L_{\gamma-1,n}^1[\gamma/(\gamma+n/2)] \quad (2.21)$$

(see (3.1) for the definition of $L_{\gamma,n}^1$).

3. Bounds for the Kinetic Energy

In this section, we shall use Theorem 1 to derive inequalities of the type (1.3). We recall the definition (2.14) and we further define

$$L_{\gamma,n}^1 \equiv \sup_V |e_1|^\gamma / \int |V|_+^{\gamma+n/2}. \quad (3.1)$$

Clearly,

$$L_{\gamma,n}^1 \leq L_{\gamma,n}. \quad (3.2)$$

If $\psi \in \mathcal{H}_{N,n,q}$ = the N -fold antisymmetric tensor product of $L^2(\mathbb{R}^n; \mathbb{C}^q)$, we can write ψ pointwise as $\psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)$ with $x_j \in \mathbb{R}^n$, $\sigma_j \in \{1, 2, \dots, q\}$ and $\psi \rightarrow -\psi$ if (x_i, σ_i) is permuted with (x_j, σ_j) . $q = 2$ for spin $1/2$ fermions. We can extend the definition (1.2) to

$$\rho_\sigma(x) \equiv N \sum_{\sigma_2=1}^q \cdots \sum_{\sigma_N=1}^q \int |\psi(x, x_2, \dots, x_N; \sigma, \sigma_2, \dots, \sigma_N)|^2 d^n x_2 \cdots d^n x_N. \quad (3.3)$$

We also define

$$T_{\psi} \equiv \sum_{j=1}^N \sum_{\sigma_1=1}^q \cdots \sum_{\sigma_N=1}^q \int |\nabla_j \psi(\underline{x}; \underline{\sigma})|^2 d^{nN} \underline{x}, \quad (3.4)$$

$$\|\psi\|_2^2 \equiv \sum_{\sigma_1=1}^q \cdots \sum_{\sigma_N=1}^q \int |\psi(\underline{x}; \underline{\sigma})|^2 d^{nN} \underline{x}. \quad (3.5)$$

Our result is

THEOREM 4. *Let p satisfy $\max\{n/2, 1\} \leq p \leq 1 + n/2$ and suppose that $L_{p-n/2, n} < \infty$. If $\|\psi\|_2 = 1$, then, except for the case $n = 2$, $p = 1$, there exists a positive constant $K_{p,n}$ such that*

$$T_{\psi} \geq K_{p,n} \sum_{\sigma=1}^q \left[\int \rho_{\sigma}(\underline{x})^{p/(p-1)} d^n \underline{x} \right]^{2(p-1)/n} \quad (3.6)$$

and

$$K_{p,n} \geq \frac{1}{2} n p^{-2p/n} (p-n/2)^{-1+2p/n} (L_{p-n/2, n}^1 / L_{p-n/2, n})^{-1+2p/n} L_{p-n/2, n}^1^{-2/n}. \quad (3.7)$$

Before giving the proof of Theorem 4, we discuss its relation to the well-known Sobolev inequalities [9, 10]:

THEOREM 5 (Sobolev-Talenti-Aubin). *Let $\nabla \psi \in L^r(\mathbb{R}^n)$ with $1 < r < n$. Let $t = nr/(n-r)$. Then*

$$\int |\nabla \psi|^r \geq C_{r,n} \left(\int |\psi|^t \right)^{r/t} \quad (3.8)$$

for some $C_{r,n} > 0$.

Talenti [11] and Aubin [21] have given the best possible $C_{r,n}$ (for $n = 3$, $r = 2$, $t = 6$, $C_{2,n}$ is also given in [8] and [12]):

$$C_{r,n} = n\pi^{r/2} \left(\frac{r-1}{n-r}\right)^{1-r} \left\{ \frac{\Gamma(1+n-n/r)\Gamma(n/r)}{\Gamma(n)\Gamma(1+n/2)} \right\}^{r/n}. \quad (3.9)$$

Our inequality (3.6) relates only to the $r = 2$ case in (3.8), in which case $t = 2n/(n-2)$. Consider (3.8) with $r = 2$ and $\|\psi\|_2 = 1$. Using Hölder's inequality on the right side of (3.8), one gets

$$\int |\nabla \psi|^2 \geq C_{2,n} \left[\int |\psi|^{2p/(p-1)} \right]^{2(p-1)/n} \left[\int |\psi|^2 \right]^{-2(p-n/2)/n} \quad (3.10)$$

whenever $n > 2$ and $p \geq n/2$. However, $C_{2,n}$ is not necessarily the best constant in (3.10) when $p \neq n/2$ ($p = n/2$ corresponds to $r = 2$ in (3.8)). Indeed, Theorem 4 says something about this question.

In the case that $N = 1$ and $q = 1$, Theorem 4 is of the same form as (3.10) (since $\rho = |\psi|^2$ and $\|\psi\|_2 = 1$). We note two things:

1. For $n > 2$ and $p = n/2$, (3.6) agrees with (3.8) except, possibly, for a different constant. We have, therefore, an alternative proof of the usual Sobolev inequality (for the $r = 2$ case). As we shall also show $K_{n/2,n} = C_{2,n}$, so we also have the best possible constant for this case.

2. If $\max\{n/2, 1\} < p \leq 1 + n/2$, Theorem 4 gives an improved version of (3.10), even if $n = 1$ or 2 (in which cases $C_{2,n} = 0$, but $K_{p,n} > 0$). For $p > 1 + n/2$, one can always use Hölder's inequality on the $p = 1 + n/2$ result to get a nontrivial bound of the form (3.10). However, in Theorem 4, the restriction $p \leq 1 + n/2$ is really necessary. This has to do with the dependence of T_ψ on N rather than on n , as we shall explain shortly.

Next we turn to the case $N > 1$. To illustrate the nature of (3.6), we may as well suppose $q = 1$. To fix ideas, we take a special, but important form for ψ , namely

$$\psi(x_1, \dots, x_N) = (N!)^{-1/2} \text{Det} \{\phi^i(x_j)\}_{ij=1}^N \quad (3.11)$$

and where the ϕ^i are orthonormal functions in $L^2(\mathbb{R}^n)$. Then, suppressing the subscript σ because $q = 1$,

$$\rho(x) = \sum_{i=1}^N \rho^i(x) ,$$

$$\rho^i(x) = |\phi^i(x)|^2 ,$$

$$T_{\psi} = \sum_{i=1}^N t^i ,$$

$$t^i = \int |\nabla \phi^i|^2 . \quad (3.12)$$

Theorem 4 says that

$$\sum_i t^i \geq K_{p,n} \left\{ \int \left[\sum_i \rho^i(x) \right]^{p/(p-1)} d^n x \right\}^{2(p-1)/n} . \quad (3.13)$$

If we did not use the orthogonality of the ϕ^i , all we would be able to conclude, using (3.6) with $N = 1$, N times, would be

$$\sum_i t^i \geq K_{p,n} \sum_i \left[\int \rho^i(x)^{p/(p-1)} d^n x \right]^{2(p-1)/n} . \quad (3.14)$$

If $p = n/2$, then (3.14) is better than (3.13), by convexity. In the opposite case, $p = 1 + n/2$, (3.13) is superior. For in between cases, (3.13) is decidedly better if N is large and if the ρ^i are close to each other (in the $L^{p/(p-1)}(\mathbb{R}^n)$ sense). Suppose $\rho^i(x) = \rho(x)/N$, $i = 1, \dots, N$. Then the right side of (3.13) is proportional to $N^{2p/n}$ while the right side of (3.14) grows only as N . This difference is caused by the orthogonality of the ϕ^i , or the Pauli principle.

In fact, the last remark shows why $p \leq 1 + n/2$ is important in Theorem 4. If $\rho^i = \rho/N$, all i , then the best bound, insofar as the N dependence is concerned, occurs when p is as large as possible. It is easy to see

by example, however, that the largest growth for T_ψ due to the orthogonality condition can only be $N^{(n+2)/n}$.

PROOF OF THEOREM 4. Let $V(x) \leq 0$ be a potential in R^n with at least one bound state. If $e_1 = \min \{e_j\}$, then, for $\gamma \in [0, 1]$,

$$\sum_j |e_j|^\gamma \geq |e_1|^{\gamma-1} \sum_j |e_j|.$$

Using the definition (2.14) and (3.2), we have that

$$\sum_j |e_j| \leq A_{\gamma,n} \left\{ \int |V|^{\gamma+n/2} \right\}^{1/\gamma} \quad (3.15)$$

$$A_{\gamma,n} = L_{\gamma,n} (L_{\gamma,n}^1)^{-1+1/\gamma} \quad (3.16)$$

when $1 \geq \gamma \geq \max(0, 1 - n/2)$. (3.15) holds even if V has no bound state.

Let π_σ , $\sigma = 1, \dots, q$, be the projection onto the state σ , i.e. for $\psi \in L^2(R^n; C^q)$, $(\pi_\nu \psi)(x, \sigma) = \psi(x, \nu)$ if $\sigma = \nu$ and zero otherwise. Choose $\gamma = p - n/2$. Let $\{\rho_\sigma\}_{\sigma=1}^q$ be given by (3.3) and, for $\alpha_\sigma \geq 0$, $\sigma = 1, \dots, q$, define

$$h = -\Delta - \sum_{\sigma=1}^q \alpha_\sigma \rho_\sigma(x)^{1/(\gamma+n/2-1)} \pi_\sigma \quad (3.17)$$

to be an operator on $L^2(R^n; C^q)$ in the usual way. Define

$$H_N = \sum_{i=1}^N h_i \quad (3.18)$$

where h_i means h acting on the i -th component of $\mathcal{H}_{N,n,q}$. Finally, let $E = \inf \text{spec } H_N$.

Now, by the Rayleigh-Ritz variational principle

$$E \leq (\psi, H_N \psi) = T_\psi - \sum_{\sigma=1}^q a_\sigma \int \rho_\sigma^{p/(p-1)}. \quad (3.19)$$

On the other hand, $E \geq$ the sum of all the negative eigenvalues of h

$$\geq -A_{\gamma,n} \sum_{\sigma=1}^q a_\sigma^{p/\gamma} \left\{ \int \rho_\sigma^{p/(p-1)} \right\}^{1/\gamma} \quad (3.20)$$

by (3.15). Combining (3.19) and (3.20) with

$$a_\sigma = \left\{ \int \rho_\sigma^{p/(p-1)} \right\}^{2(\gamma-1)/n} \left\{ \frac{\gamma}{\gamma + n/2} \frac{1}{A_{\gamma,n}} \right\}^{2\gamma/n},$$

the theorem is proved.

Note that when $p = 1 + n/2$ (corresponding to $\gamma = 1$ in the proof), $L_{1,n}^1$ does not appear in (3.7). In this case, the right side of (3.7) is the best possible value of $K_{1+n/2,n}$, as we now show.

LEMMA 6. From (3.7), define

$$L_{1,n}^* \equiv [n/(2K_{1+n/2,n})]^{n/2} (1 + n/2)^{-1-n/2}.$$

Then $L_{1,n}^* = L_{1,n}$.

PROOF. By (3.7), we only have to prove that $L_{1,n}^* \geq L_{1,n}$. Let $V \leq 0$, $V \in C_0^\infty(\mathbb{R}^n)$ and let $H = -\Delta + V$. Let $\{\phi_i, e_i\}_{i=1}^N$ be the bound state eigenfunctions and eigenvalues of H . Let ψ and ρ^i be as defined in (3.11), (3.12). Then

$$\sum_i |e_i| = - \int V \rho - T_\psi \leq \|V\|_p \|\rho\|_{p/(p-1)} - T_\psi$$

with $p = 1 + n/2$. Using Theorem 4 for T_ψ , one has that

$$\sum_i |e_i| \leq \max_{y>0} \{ \|V\|_p y - K_{1+n/2,n} y^{2p/n} \} = L_{1,n}^* \|V\|_p^p .$$

We conclude with an evaluation of $K_{n/2,n}$ for $n > 2$ as promised. By a simple limiting argument

$$K_{n/2,n} \geq \lim_{p \downarrow n/2} (\text{right side of (3.7)}) . \quad (3.21)$$

Our bound (2.11) on $L_{p-n/2,n}$ shows that

$$\lim_{p \downarrow n/2} (L_{p-n/2,n})^{-1+2p/n} = 1 . \quad (3.22)$$

Hence

$$K_{n/2,n} \geq (L_{0,n}^1)^{-2/n} . \quad (3.23)$$

On the other hand, by the method of Lemma 6 applied to the $N = 1$ case, $K_{n/2,n} \leq (L_{0,n}^1)^{-2/n}$. The value of $L_{0,n}^1$ is given in (4.24). To be honest, its evaluation requires the solution of the same variational problem as given in [8, 11, 12]. Substitution of (4.24) into (3.23) yields the required result

$$K_{n/2,n} = C_{2,n} = \pi n(n-2) [\Gamma(n/2)/\Gamma(n)]^{2/n} . \quad (3.24)$$

If we examine (3.23) when $n = 2$, one gets $K_{1,2} \geq 0$ since $L_{0,2}^1 = \infty$. This reflects the known fact [5] that an arbitrarily small $V < 0$ always has a bound state in two dimensions. This observation can be used to show that

$$K_{1,2} = 0 . \quad (3.25)$$

When $n = 1$, the smallest allowed p is $p = 1$. In this case, (3.6) reads

$$T_\psi \geq K_{1,1} \sum_{\sigma=1}^q \|\rho_\sigma\|_\infty^2 . \quad (3.26)$$

Using (3.7) and (4.20),

$$K_{1,1} \geq [2L_{1/2,1}]^{-1} . \quad (3.27)$$

If one accepts the conjecture of Section 4 that $L_{1/2,1} = L_{1/2,1}^1 = 1/2$, then

$$K_{1,1} = 1 . \quad (3.28)$$

The reason for the equality in (3.28) is that $K_{1,1} = 1$ is well known to be the best possible constant in (3.26) when $q = 1$ and $N = 1$.

4. Conjecture About $L_{\gamma,n}$

We have shown that for the bound state energies $\{e_j\}$ of a potential V in n dimensions and with

$$L_{\gamma,n}(V) \equiv \sum_j |e_j|^\gamma / \int |V|^{\gamma+n/2} , \quad (4.1)$$

then

$$L_{\gamma,n} \equiv \sup_{V \in L^{\gamma+n/2}} L_{\gamma,n}(V) \quad (4.2)$$

is finite whenever $\gamma + n/2 > 1$ and $\gamma > 0$. The “boundary points” are

$$\begin{array}{ll} \gamma = 1/2 & n = 1 \\ \gamma = 0 & n \geq 2 . \end{array} \quad (4.3)$$

We showed that for $n = 1$, $L_{1/2,1} < \infty$. For $\gamma < 1/2$, $n = 1$, there cannot be a bound of this kind, for consider $V_L(x) \equiv -1/L$ for $|x| < L$ and zero otherwise. For $L \rightarrow 0$, this converges towards $-2\delta(x)$ and thus has a bound state of finite energy (which is -1 for $-2\delta(x)$). On the other hand,

$$\lim_{L \rightarrow 0} \int dx |V_L|^{1/2+\gamma} = 0 \quad \text{for } \gamma < 1/2 .$$

For $n = 2$, $\gamma = 0$ is a “double boundary point” and $L_{0,2} = \infty$, i.e. there is no upper bound on the number of bound states in two dimensions. (Cf. [5].)

For $n \geq 3$, $L_{0,n}$ is *conjectured* to be finite (see note added in proof, Section 5); for $n = 3$, this is the well-known $\int |V|_-^{3/2}$ conjecture on the number, $N_0(V)$, of bound states (cf. [5]). The best that is known at present is that

$$N_0(V) \leq c \left[\int |V|_-^{3/2} \right]^{4/3}, \quad (4.4)$$

but for spherically symmetric V , a stronger result is known [8]:

$$N_0(V) \leq I \left(1 + \frac{1}{4} \ln I \right),$$

$$I = 4(3\pi^2)^{-1} \int |V|_-^{3/2}. \quad (4.5)$$

In (1.4) and (3.1), we introduced L^C and L^1 and showed that

$$L_{\gamma,n} \geq \max(L_{\gamma,n}^1, L_{\gamma,n}^C). \quad (4.6)$$

A parallel result is Simon's [22] for $n \geq 3$:

$$N_0(V) \leq D_{n,\epsilon} (\|V_-\|_{\epsilon+n/2} + \|V_-\|_{-\epsilon+n/2})^{n/2}$$

with $D_{n,\epsilon} \rightarrow \infty$ as $\epsilon \rightarrow 0$.

In our previous paper [4], we conjectured that $L_{1,3} = L_{1,3}^C$, and we also pointed out that $L_{1,1}^1 > L_{1,1}^C$. A remark of Peter Lax (private communication), which will be explained presently, led us to the following:

CONJECTURE. *For each n , there is a critical value of $\gamma, \gamma_{C,n}$, such that*

$$L_{\gamma,n} = L_{\gamma,n}^C \quad \gamma \geq \gamma_C$$

$$L_{\gamma,n} = L_{\gamma,n}^1 \quad \gamma \leq \gamma_C$$

γ_c is defined to be that γ for which $L_{\gamma,n}^C = L_{\gamma,n}^1$; the uniqueness of this γ_c is part of the conjecture. Furthermore, $\gamma_{c,1} = 3/2$, $\gamma_{c,2} \sim 1.2$, $\gamma_{c,3} \sim .86$ and the smallest n such that $\gamma_{c,n} = 0$ is $n = 8$.

(A) Remarks on $L_{\gamma,n}^1$

We want to maximize

$$\left| \int [|\psi|^2 V + |\nabla \psi|^2] d^n x \right|^\gamma / \int |V|^{\gamma+n/2} \quad (4.7)$$

with respect to V , and where $\int |\psi|^2 = 1$ and $(-\Delta + V)\psi = e_1 \psi$. By the variational principle, we can first maximize (4.7) with respect to V , holding ψ fixed. Hölder's inequality immediately yields

$$V(x) = -\sigma |\psi(x)|^{2/(\gamma+n/2-1)}$$

with $\sigma > 0$. The kinetic energy, $\int |\nabla \psi|^2$, is not increased if $\psi(x)$ is replaced by $|\psi(x)|$ and, by the rearrangement inequality [13], this is not increased if $|\psi|$ is replaced by its symmetric decreasing rearrangement. Thus, we may assume that $|V|$ and $|\psi|$ are spherically symmetric, non-increasing functions.

By the methods of [8] or [11], (4.7) can be shown to have a maximum when $\gamma + n/2 > 1$. The variational equation is

$$-\Delta \psi(x) - \sigma \psi(x)^{(\gamma+n/2+1)/(\gamma+n/2-1)} = e_1 \psi(x) \quad (4.8)$$

with

$$\sigma = \left\{ \frac{\gamma |e_1|^{\gamma-1}}{(\gamma + n/2) L_{\gamma,n}^1} \right\}^{1/(\gamma+n/2-1)}. \quad (4.9)$$

Equation (4.8) determines ψ up to a constant and up to a change of scale in x . The former can be used to make $\int \psi^2 = 1$ and the latter leaves (4.7) invariant.

Equation (4.8) can be solved analytically in two cases, to which we shall return later:

- (i) $n = 1$, all $\gamma > 1/2$
- (ii) $n \geq 3$, $\gamma = 0$.

(B) The One-Dimensional Case

Lax's remark was about a result of Gardner, Greene, Kruskal and Miura [14] to the effect that

$$L_{3/2,1} = L_{3/2,1}^C = 3/16. \quad (4.10)$$

To see this, we may assume $V \in C_0^\infty(\mathbb{R})$, and use the theory of the Korteweg-de Vries (KdV) equation [14]:

$$W_t = 6WW_x - W_{xxx}. \quad (4.11)$$

There are two remarkable properties of (4.11):

(i) As W evolves in time, t , the eigenvalues of $-d^2/dx^2 + W$ remain invariant.

(ii) $\int W^2 dx$ is constant in time.

Let $W(x, t)$ be given by (4.11) with the initial data

$$W(x, 0) = V(x).$$

Then $L_{3/2,1}(W(\cdot, t))$ is independent of t , and may therefore be evaluated by studying its behavior as $t \rightarrow \infty$.

There exist traveling wave solutions to (4.11), called solitons, of the form

$$W(x, t) = f(x - ct).$$

Equation (4.11) becomes

$$-cf_x = -f_{xxx} + 6ff_x. \quad (4.12)$$

The solutions to (4.12) which vanish at ∞ are

$$\begin{aligned} f_a(x) &= -2a^2 \cosh^{-2}(ax) \\ c &= 4a^2. \end{aligned} \quad (4.13)$$

Any solution (4.13), regarded as a potential in the Schrödinger equation has, as we shall see shortly, exactly one negative energy bound state with energy and wave function

$$e = -a^2$$

$$\psi(x) = \cosh^{-1}(ax) . \quad (4.14)$$

Now the theory of the KdV equation says that as $t \rightarrow \infty$, W evolves into a sum of solitons (4.13) plus a part that goes to zero in $L^\infty(\mathbb{R})$ norm (but not necessarily in $L^2(\mathbb{R})$ norm). The solitons are well separated since they have different velocities. Because the number of bound states is finite, the non-soliton part of W can be ignored as $t \rightarrow \infty$. Hence, for the initial V ,

$$\sum |e_j|^{3/2} = \sum_{\text{solitons}} a^3 \quad (4.15)$$

while

$$\int V(x)^2 dx \geq \sum_{\text{solitons}} \int f_a(x)^2 dx . \quad (4.16)$$

Since $4 \int_{-\infty}^{\infty} \cosh^{-4}(x) dx = 16/3$, we conclude that

$$L_{3/2,1} = L_{3/2,1}^C = 3/16 \quad (4.17)$$

with equality if and only if $W(x, t)$ is composed purely of solitons as $t \rightarrow \infty$. For the same reason,

$$L_{3/2,1}^1 = L_{3/2,1}^C \quad (4.18)$$

(cf. (4.21)).

Not only do we have an evaluation of $L_{3/2,1}$, (4.17), but we learn something more. When $\gamma = 3/2$, there is an infinite family of potentials for which $L_{3/2,1}(V) = L_{3/2,1}$, and these may have any number of bound states = number of solitons.

What we believe to be the case is that when $\gamma < 3/2$, the optimizing potential for $L_{\gamma,n}$ has only one bound state, and satisfies (4.8). When

$\gamma > 3/2$, the optimizing potential is, loosely speaking, infinitely deep and has infinitely many bound states; thus $L_{\gamma,n} = L_{\gamma,n}^C$.

An additional indication that the conjecture is correct is furnished by the solution to (4.8). When $\gamma = 3/2$, this agrees with (4.14). In general, one finds that, apart from scaling, the nodeless solution to (4.8) is

$$\begin{aligned}\psi_{\gamma}(x) &= \Gamma(\gamma)^{1/2} \pi^{-1/4} \Gamma(\gamma - 1/2)^{-1/2} \cosh^{-\gamma+1/2}(x) \\ V_{\gamma}(x) &= -(\gamma^2 - 1/4) \cosh^{-2}(x) \\ e_1 &= -(\gamma - 1/2)^2 .\end{aligned}\tag{4.19}$$

Thus,

$$L_{\gamma,1}^1 = \pi^{-1/2} \frac{1}{\gamma - 1/2} \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1/2)} \left(\frac{\gamma-1/2}{\gamma+1/2} \right)^{\gamma+1/2} .\tag{4.20}$$

When $L_{\gamma,1}^1$ is compared with $L_{\gamma,1}^C$, one finds that

$$\begin{aligned}L_{\gamma,1}^1 &\geq L_{\gamma,1}^C & \gamma &\leq 3/2 \\ L_{\gamma,1}^1 &\leq L_{\gamma,1}^C & \gamma &\geq 3/2 .\end{aligned}\tag{4.21}$$

This confirms at least part of the conjecture.

However, more is true. For $\gamma = 3/2$, V_{γ} has a zero energy single node bound state

$$\phi(x) = \tanh(x) .$$

Since V_{γ} is monotone in γ , it follows that V_{γ} has only one bound state for $\gamma < 3/2$ and at least two bound states for $\gamma > 3/2$. The (unnormalized) second bound state can be computed to be

$$\begin{aligned}\phi(x) &= \sinh(x) \cosh^{-\gamma+1/2}(x) \\ e_2 &= -(\gamma - 3/2)^2 .\end{aligned}\tag{4.22}$$

In like manner, one can find more bound states as γ increases even further.

Thus we see that the potential that optimizes the ratio $|e_1|^\gamma / \int |V|^{\gamma+1/2}$ automatically has a second bound state when $\gamma > \gamma_c$.

Finally, we remark that Theorem 3, together with (4.10), shows that

THEOREM 7.

$$L_{\gamma,1} = L_{\gamma,1}^C \quad \text{for} \quad \gamma = 3/2, 5/2, 7/2, \text{ etc.}$$

An application of Theorem 7 to scattering theory will be made in Section 4(D).

(C) Higher Dimensions

We have exhibited the solution to the variational equation (4.8) for $L_{\gamma,1}^1$.

When $n > 2$ and $\gamma = 0$, we clearly want to take $e_1 = 0$ in order to maximize $L_{0,n}^1(V)$. (4.8) has the zero energy solution

$$\phi(x) = (1 + |x|^2)^{1-n/2}$$

$$V(x) = \alpha \phi(x)^{2/(\gamma+n/2-1)} = n(n-2)(1 + |x|^2)^{(2-n)/(n/2-1)} \quad (4.23)$$

(note: $\phi \in L^2(\mathbb{R}^n)$ if and only if $n > 4$, but $V \in L^{n/2}(\mathbb{R}^n)$ always). This leads to

$$L_{0,n}^1 = [\pi n(n-2)]^{-n/2} \Gamma(n)/\Gamma(n/2) . \quad (4.24)$$

The smallest dimension for which $L_{0,n}^1 \leq L_{0,n}^C$ is $n = 8$.

If we suppose that the ratio $L_{\gamma,n}^1/L_{\gamma,n}^C$ is monotone decreasing in γ (as it is when $n = 1$ and as it is when $n = 3$ on the basis of the numerical solution of (4.10) by J. F. Barnes, given in Appendix A), and if our conjecture is correct, then $L_{\gamma,n} = L_{\gamma,n}^C$ for $n \geq 8$. The value of γ_c obtained numerically is

$$\begin{aligned} \gamma_c &= 1.165 & n &= 2 \\ \gamma_c &= .863 & n &= 3 . \end{aligned} \quad (4.26)$$

The other bit of evidence, apart from the monotonicity of $L^1_{\gamma,n}/L^C_{\gamma,n}$, for the correctness of our conjecture is a numerical study of the energy levels of the potential

$$V_\lambda(x) = \lambda e^{-|x|}, \qquad \lambda > 0,$$

in three dimensions. This is given in Appendix A. The energy levels of the square well potential are given in [15, 16]. In both cases, one finds that

$$\lim_{\lambda \rightarrow \infty} L_{1,3}(V_\lambda) = L^C_{1,3}$$

and the limit is approached from below. Unfortunately, it is not true, as one might have hoped, that $L_{1,3}(V_\lambda)$ is monotone increasing in λ .

(D) Bounds on One-Dimensional Scattering Cross-Sections

In their study of the KdV equation, (4.11), Zakharov and Fadeev [17] showed how to relate the solution $W(x, t)$ to the scattering reflection coefficient $R(k)$ and the bound state eigenvalues $\{e_j\}$ of the initial potential $V(x)$. There are infinitely many invariants of (4.11) besides $\int W^2$ and these have simple expressions in terms of $R(k)$, $\{e_j\}$.

Thus, for any potential V ,

$$\int V^2 = (16/3) \sum |e_j|^{3/2} - 4 \int_{-\infty}^\infty k^2 T(k) dk \qquad (4.27)$$

$$\int V^3 + \frac{1}{2} V_x^2 = -(32/5) \sum |e_j|^{5/2} - 8 \int_{-\infty}^\infty k^4 T(k) dk \qquad (4.28)$$

$$\int V^4 + 2VV_x^2 + \frac{1}{5} V_{xx}^2 = (256/35) \sum |e_j|^{7/2} - (64/5) \int_{-\infty}^\infty k^6 T(k) dk \qquad (4.29)$$

where

$$T(k) = \pi^{-1} \ln(1 - |R(k)|^2) \leq 0. \qquad (4.30)$$

These are only the first three invariants; a recursion relation for the others can be found in [17].

Notice that $3/16$, $5/32$, $35/256$ are, respectively $L_{3/2,1}^C$, $L_{5/2,1}^C$, $L_{7/2,1}^C$. Since $\int V^2 \geq \int |V_-|^2$, (4.27) establishes that $L_{3/2,1} = L_{3/2,1}^C$, as mentioned earlier. For the higher invariants, the signs in (4.28) and (4.29) are not as fortunately disposed and we cannot use these equations to prove Theorem 7. But, given that Theorem 7 has already been proved, we can conclude that

THEOREM 8. *For any nonpositive potential $V(x)$,*

$$\int V_x^2 \geq -16 \int_{-\infty}^{\infty} k^4 T(k) dk . \quad (4.31)$$

For any potential $V(x)$,

$$2 \int V V_x^2 + (1/5) \int V_{xx}^2 \leq -(64/5) \int_{-\infty}^{\infty} k^6 T(k) dk . \quad (4.32)$$

The first inequality, (4.31), is especially transparent: If $V(x)$ is very smooth, it cannot scatter very much.

5. Note Added in Proof

After this paper was written, M. Cwikel and Lieb, simultaneously and by completely different methods, showed that the number of bound states, $N_0(V)$ for a potential, V , can be bounded (when $n \geq 3$) by

$$N_0(V) \leq A_n \int |V(x)|_-^{n/2} d^n x . \quad (5.1)$$

Cwikel exploits the weak trace ideal method of Simon [22]; his method is more general than Lieb's, but for the particular problem at hand, (5.1), his

A_n does not seem to be as good. Lieb's method uses Wiener integrals and the general result is the following:

$$N_{-\alpha}(V) \leq \int d^n x \int_0^\infty dt t^{-1} e^{-\alpha t} (4\pi t)^{-n/2} f(t|V(x)|_-) \quad (5.2)$$

for any non-negative, convex function $f: [0, \infty) \rightarrow [0, \infty)$ satisfying

$$1 = \int_0^\infty t^{-1} f(t) e^{-t} dt. \quad (5.3)$$

For $\alpha = 0$, one can choose $f(t) = c(t-b)$, $t \geq b$, $f(t) = 0$, $t \leq b$. This leads to (5.1), and optimizing with respect to b , one finds that

$$A_3 = 0.116, \quad A_4 = 0.0191 \quad (5.4)$$

and, as $n \rightarrow \infty$,

$$A_n/L_{0,n}^C = (n\pi)^{1/2} + O(n^{-1/2}). \quad (5.5)$$

Note that $A_3/L_{0,3}^1 = 1.49$; i.e. A_3 exceeds $L_{0,3}$ by at most 49%.

Since $N_{-\alpha}(V) \leq N_0(-|V+\alpha|_-)$, one can use (5.1) and (2.1) to deduce that for $\gamma \geq 0$ and $n \geq 3$,

$$L_{\gamma,n} \leq L_{\gamma,n}^C(A_n/L_{0,n}^C). \quad (5.6)$$

This is better than (2.11), (2.18). In particular, for $n = 3$, $\gamma = 1$, the improvement of (5.6) over (2.11) with $m = 2$ is a factor of 1.83. The factor 1.31 in Equation (1.4) can therefore be replaced by $1.31 (1.83)^{-2/3} = 0.87$.

APPENDIX A. NUMERICAL STUDIES

John F. Barnes
Theoretical Division
Los Alamos Scientific Laboratory
Los Alamos, New Mexico 87545

 I. *Evaluation of $L_{\gamma,n}^1$, $n = 1, 2, 3$*

The figure shows the numerical evaluation of $L_{\gamma,n}^1$ as well as $L_{\gamma,n}^C$. The latter is given in (1.11)

$$L_{\gamma,n}^C = 2^{-n} \pi^{-n/2} \Gamma(\gamma+1) / \Gamma(\gamma+1+n/2) .$$

The former is obtained by solving the differential equation (4.8) in polar coordinates and choosing σ such that $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$. Note that by scaling, one can take $e_1 = -1$, whence

$$(L_{\gamma,n}^1)^{-1} = \sigma^{(\gamma+n/2)} \int |\psi(x)|^{(2\gamma+n)/(\gamma-1+n/2)} d^n x .$$

In one dimension, $L_{\gamma,1}^1$ is known analytically and is given in (4.20). Another exact result, (4.24), is

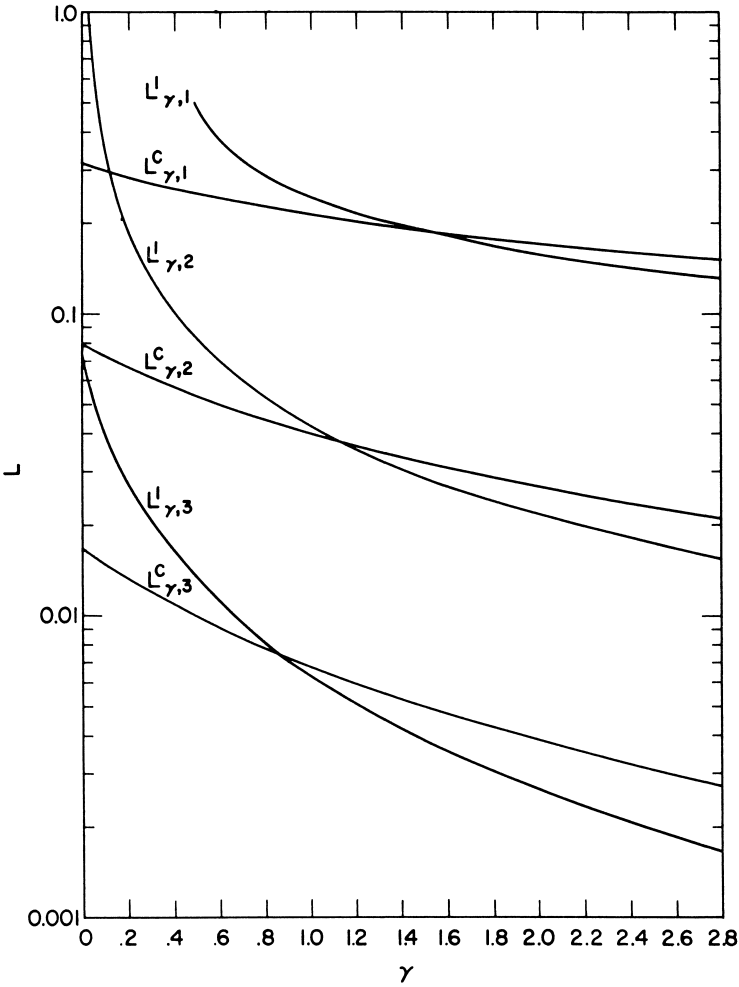
$$L_{0,3}^1 = 4\pi^{-2} 3^{-3/2} = 0.077997 .$$

The critical values of γ , at which $L_{\gamma,n}^1 = L_{\gamma,n}^C$ are:

$$\gamma_{c,1} = 3/2$$

$$\gamma_{c,2} = 1.165$$

$$\gamma_{c,3} = 0.8627 .$$



II. The Exponential Potential

To test the conjecture that $L_{1,3} = L_{1,3}^C$, the eigenvalues of the potential $V_\lambda = -\lambda \exp(-|x|)$ in three dimensions were evaluated for $\lambda = 5, 10, 20, 30, 40, 50$, and 100 . These are listed in the table according to angular momentum and radial nodes. These numbers have been corroborated by H. Grosse, and they can be used to calculate $L_{\gamma,3}(V_\lambda)$ for any γ . The final column gives $L_{1,3}(V_\lambda)$, since $\int |V_\lambda|^{5/2} = \lambda^{5/2} (64\pi)/125$. It is to be noted that the classical value $L_{1,3}^C = 0.006755$, is approached from below, in agreement with the conjecture, but *not monotonically*.

$V_\lambda = -\lambda e^{-r}$						
	ℓ	$ e $	nodes	states	$\sum e $	$\frac{\sum e }{\lambda^{5/2} \frac{64\pi}{125}}$
	—	—	—	—	—	—
$\lambda = 5$	0	0.55032	0	$\frac{1}{1}$	$\frac{0.55032}{0.55032}$	0.006120
$\lambda = 10$	0	0.06963	1			0.006398
		2.18241	0	2	2.2520	
	1	0.33405	0	$\frac{3}{5}$	$\frac{1.0022}{3.2542}$	
$\lambda = 20$	0	0.00869	2			0.006551
		1.42562	1			
		6.62410	0	3	8.0584	
	1	0.16327	1			
		2.71482	0	6	8.6342	
	2	0.43136	0	$\frac{5}{14}$	$\frac{2.1568}{18.8494}$	

$V_\lambda = -\lambda e^{-r}$ (continued)						$\frac{\sum e }{\lambda^{5/2} \frac{64\pi}{125}}$
	ℓ	$ e $	nodes	states	$\sum e $	
$\lambda = 30$	0	0.58894	2			
		3.83072	1			
		11.84999	0	3	16.270	
	1	1.39458	1			
		6.12302	0	6	22.553	
	2	0.00593	1			
		2.36912	0	10	11.875	
	3	0.07595	0	$\frac{7}{26}$	$\frac{0.532}{51.230}$	0.006461
$\lambda = 40$	0	0.07676	3			
		1.86961	2			
		6.88198	1			
		17.53345	0	4	26.362	
	1	0.41991	2			
		3.35027	1			
		10.13596	0	9	41.718	
	2	0.93459	1			
		5.03378	0	10	29.842	
	3	1.54738	0	$\frac{7}{30}$	$\frac{10.832}{108.754}$	0.006682

$V_\lambda = -\lambda e^{-r}$ (continued)						$\frac{\sum e }{\lambda^{5/2} \frac{64\pi}{125}}$
	ℓ	$ e $	nodes	states	$\sum e $	
$\lambda = 50$	0	0.60190	3			
		3.66447	2			
		10.39110	1			
		23.53215	0	4	38.190	
	1	1.43321	2			
		5.81695	1			
		14.56904	0	9	65.458	
	2	0.07675	2			
		2.45887	1			
		8.19840	0	15	53.670	
	3	0.26483	1			
		3.61626	0	14	27.168	
	4	0.49009	0	$\frac{9}{51}$	$\frac{4.411}{188.897}$	0.006643
$\lambda = 100$	0	0.39275	5			
		2.91408	4			
		8.29231	3			
		17.44909	2			
		32.07168	1			

$V_\lambda = -\lambda e^{-r}$ (continued)					
ℓ	$ e $	nodes	states	$\sum e $	$\frac{\sum e }{\lambda^{5/2} \frac{64\pi}{125}}$
1	56.28824	0	6	117.41	
	1.10170	4			
	4.76748	3			
	11.62740	2			
	22.79910	1			
2	40.45495	0	15	242.25	
	0.02748	4			
	2.04022	3			
	6.85633	2			
	15.22147	1			
3	28.46495	0	25	263.05	
	0.22692	3			
	3.14743	2			
	9.13429	1			
4	19.04073	0	28	220.85	
	0.52962	2			
	4.37856	1			
	11.56470	0	27	148.26	

$$V_\lambda = -\lambda e^{-r} \quad (\text{continued})$$

ℓ	$ e $	nodes	states	$\sum e $	$\frac{\sum e }{\lambda^{5/2} \frac{64\pi}{125}}$
5	0.88997	1			
	5.69707	0	22	72.46	
6	1.26789	0	$\frac{13}{136}$	$\frac{16.48}{1080.76}$	0.006719

APPENDIX B: PROOF OF (2.3)

THEOREM 9. Let $\mathcal{H}$ be a separable Hilbert space and let A, B be positive operators on $\mathcal{H}$. Then, for $m \geq 1$,

$$\text{Tr}(B^{1/2} A B^{1/2})^m \leq \text{Tr} B^{m/2} A^m B^{m/2}. \quad (\text{B.1})$$

REMARK. When $\mathcal{H} = L^2(\mathbb{R}^n)$ and A is a kernel $a(x-y)$ and B is a multiplication operator $b(x)$ (as in our usage (2.2)), Seiler and Simon [19] have given a proof of (B.1) using interpolation techniques. Simon (private communication) has extended this method to the general case. Our proof is different and shows a little more than just (B.1).

PROOF. For simplicity, we shall only give the proof when A and B are matrices; for the general case, one can appeal to a limiting argument. For $m = 1$, the theorem is trivial, so assume $m > 1$. Let $C = A^m$ and $f(C) \equiv g(C) - h(C)$, where $g(C) \equiv \text{Tr}(B^{1/2} C^{1/m} B^{1/2})^m$ and $h(C) = \text{Tr} B^{m/2} C B^{m/2}$. Let M^+ be the positive matrices. Clearly $M^+ \ni C \rightarrow h(C)$ is linear. Epstein [18] has shown that $M^+ \ni C \rightarrow g(C)$ is concave (actually, he showed this for m integral, but his proof is valid generally for $m \geq 1$). Write $C = C^D + C^0$ where C^D is the diagonal part of C in a basis in which B is diagonal. $C_\lambda \equiv C^D + \lambda C^0 = \lambda C + (1-\lambda)C^D$ is in M^+ for $\lambda \in [0, 1]$,

because $C^D \in M^+$. Then $\lambda \rightarrow f(C_\lambda) \equiv R(\lambda)$ is concave on $[0, 1]$. Our goal is to show that $R(1) \leq 0$. Since $[C^D, B] = 0$, $R(0) = 0$ and, by concavity, it is sufficient to show that $R(\lambda) \leq 0$ for $\lambda > 0$ and λ small. $h(C_\lambda) = h(C^D)$ for $\lambda \in [0, 1]$. Since $f(C)$ is continuous in C , we can assume that C^D is nondegenerate and strictly positive, and that C_λ is positive when $\lambda \geq -\epsilon$ for some $\epsilon > 0$. Then $R(\lambda)$ is defined and concave on $[-\epsilon, 1]$. $\lambda \rightarrow C_\lambda^{1/m}$ is differentiable at $\lambda = 0$ and its derivative at $\lambda = 0$ has zero diagonal elements. (To see this, use the representation $C^{1/m} = K \int_0^\infty dx x^{-1+1/m} \cdot C(C+xI)^{-1}$.) Likewise, the derivative of $(B^{1/2}(D + \lambda O)B^{1/2})^m$ at $\lambda = 0$ has zero diagonal elements when O has and when D is diagonal. Thus

$$dR(\lambda)/d\lambda|_{\lambda=0} = 0.$$

Acknowledgment

One of the authors (Walter Thirring) would like to thank the Department of Physics of the University of Princeton for its hospitality.

ELLIOTT H. LIEB
DEPARTMENTS OF MATHEMATICS AND PHYSICS
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY

WALTER E. THIRRING
INSTITUT FÜR THEORETISCHE PHYSIK
DER UNIVERSITÄT WIEN, AUSTRIA

REFERENCES

- [1] E. H. Lieb and W. E. Thirring, *Phys. Rev. Lett.* **35**, 687 (1975). See *Phys. Rev. Lett.* **35**, 1116 (1975) for errata.
- [2] M. S. Birman, *Mat. Sb.* **55**(97), 125 (1961); *Amer. Math. Soc. Translations Ser. 2*, **53**, 23 (1966).
- [3] J. Schwinger, *Proc. Nat. Acad. Sci.* **47**, 122 (1961).
- [4] B. Simon, "Quantum Mechanics for Hamiltonians Defined as Quadratic Forms," Princeton University Press, 1971.

- [5] B. Simon, "On the Number of Bound States of the Two Body Schrödinger Equation – A Review," in this volume.
- [6] A. Martin, *Helv. Phys. Acta* 45, 140 (1972).
- [7] H. Tamura, *Proc. Japan Acad.* 50, 19 (1974).
- [8] V. Glaser, A. Martin, H. Grosse and W. Thirring, "A Family of Optimal Conditions for the Absence of Bound States in a Potential," in this volume.
- [9] S. L. Sobolev, *Mat. Sb.* 46, 471 (1938), in Russian.
- [10] ———, *Applications of Functional Analysis in Mathematical Physics*, Leningrad (1950), Amer. Math. Soc. Transl. of Monographs, 7 (1963).
- [11] G. Talenti, Best Constant in Sobolev's Inequality, Istituto Matematico, Università Degli Studi Di Firenze, preprint (1975).
- [12] G. Rosen, *SIAM Jour. Appl. Math.* 21, 30 (1971).
- [13] H. J. Brascamp, E. H. Lieb and J. M. Luttinger, *Jour. Funct. Anal.* 17, 227 (1974).
- [14] C. S. Gardner, J. M. Greene, M. D. Kruskal and R. M. Miura, *Commun. Pure and Appl. Math.* 27, 97 (1974).
- [15] S. A. Moszkowski, *Phys. Rev.* 89, 474 (1953).
- [16] A. E. Green and K. Lee, *Phys. Rev.* 99, 772 (1955).
- [17] V. E. Zakharov and L. D. Fadeev, *Funkts. Anal. i Ego Pril.* 5, 18 (1971). English translation: *Funct. Anal. and its Appl.* 5, 280 (1971).
- [18] H. Epstein, *Commun. Math. Phys.* 31, 317 (1973).
- [19] E. Seiler and B. Simon, "Bounds in the Yukawa Quantum Field Theory," Princeton preprint (1975).
- [20] W. Thirring, *T7 Quantenmechanik*, Lecture Notes, Institut für Theoretische Physik, University of Vienna.
- [21] T. Aubin, *C. R. Acad. Sc. Paris* 280, 279 (1975). The results are stated here without proof; there appears to be a misprint in the expression for $C_{r,n}$.
- [22] B. Simon, "Weak Trace Ideals and the Bound States of Schrödinger Operators," Princeton preprint (1975).

ON SEMI-CLASSICAL BOUNDS FOR EIGENVALUES OF SCHRÖDINGER OPERATORS^{*}

Michael AIZENMAN

Department of Physics, Princeton University, Princeton, NJ 08540, USA

and

Elliott H. LIEB

Departments of Mathematics and Physics, Princeton University Princeton, NJ 08540, USA

Received 27 April 1978

Our principal result is that if the semiclassical estimate is a bound for some moment of the negative eigenvalues (as is known in some cases in one-dimension), then the semiclassical estimates are also bounds for all higher moments.

Bounds on the moments of energy levels of Schrödinger operators have been the object of several studies [1, 2, 5–8]. In [1] such bounds were used to obtain a lower bound for the kinetic energy of fermions in terms of their one particle density and thereby prove the stability of matter.

In the notation of [2],

$$\sum_j |e_j(V)|_-^\gamma \leq L_{\gamma,n} \int d^n x |V(x)|_-^{\gamma+n/2} \quad (1)$$

where $e_j(V)$ are the eigenvalues of $-\Delta + V(x)$ defined in $L^2(\mathbf{R}^n)$ and $|y|_- \equiv \max(-y, 0)$. $L_{\gamma,n}$ denotes the smallest number for which (1) holds independently of V . The case $\gamma = 1$ is the one needed for the kinetic energy bound. It was shown in [2] that $L_{\gamma,n} < \infty$ for $\gamma > \max(0, 1 - n/2)$. Eq. (1) also holds for $\gamma = 0$, $n \geq 3$ but the proof is quite different (see refs. [5, 6, 8]).

For $\gamma \geq 0$ we use the notation

$$H(x, p) = p^2 + V(x)$$

$$R_n(\gamma, V) = \sum_j |e_j(V)|_-^\gamma / \int d^n z |H(x, p)|_-^\gamma \quad (2)$$

$$R_n(\gamma) = \sup_V \{R_n(\gamma, V)\}, \quad (3)$$

and $d^n z = d^n x d^n p (2\pi)^{-n}$. $R_n(\gamma, V)$ is the ratio of the moments of the binding energies of a quantum mechanical hamiltonian to the moments of its classical analog. The integral in (1) comes from doing the $d^n p$ integration in (2). In the notation of [2]

$$R_n(\gamma) = L_{\gamma,n} / L_{\gamma,n}^c \quad (4)$$

For $V \leq 0$, $V \in C_0^\infty(\mathbf{R}^n)$ it is known [3,4,9] that

$$R_n(\gamma, \lambda V) \rightarrow 1 \quad (5)$$

as $\lambda \rightarrow \infty$, which is the semiclassical limit. Thus

$$R_n(\gamma) \geq 1. \quad (6)$$

In [2] it was conjectured that $R_n(\gamma) = 1$ for certain γ and n , in particular for $\gamma = 1$, $n = 3$ which is the case of primary physical interest. $R_3(1) = 1$ would imply that the Thomas–Fermi theory of atoms and molecules (together with a modified treatment of the electron-electron repulsion) gives a lower bound to the true Schrödinger ground state energy (see [1]). The only cases where the value of $R_n(\gamma)$ is known are $n = 1$, $\gamma = 3/2, 5/2, 7/2, \dots$, where $R_1(\gamma) = 1$. Part (a) of the following theorem, together with (6), settles the question for $n = 1$, $\gamma \geq 3/2$.

Theorem:

(a) For any n , $R_n(\gamma)$ is a monotone nonincreasing function of γ .

^{*} Work partly supported by U.S. National Science Foundation grant MCS 75-21684 A02.

(b) If, for some $\gamma > \max(0, 1 - n/2)$, the supremum in (3) is attained, i.e. $R_n(\gamma) = R_n(\gamma, V)$ for some V with $|V|_- \in L^{\gamma+n/2}$, then $R_n(\cdot)$ is strictly decreasing from the left at γ . In fact

$$\liminf_{\delta \rightarrow 0^+} [R_n(\gamma - \delta) - R_n(\gamma)] / \delta > 0.$$

Proof:

(a) Fix V . For $\gamma \geq 0$, $\delta > 0$, let

$$I(\gamma, \delta) = \int_0^1 d\lambda \lambda^{-1+\delta} |1 - \lambda|_-^\gamma < \infty. \quad (7)$$

By scaling, for any $e \in R$

$$|e|_-^{\gamma+\delta} = I(\gamma, \delta)^{-1} \int_0^\infty d\lambda \lambda^{-1+\delta} |e + \lambda|_-^\gamma. \quad (8)$$

Thus, for any Schrödinger potential V ,

$$\begin{aligned} \sum_j |e_j(V)|_-^{\gamma+\delta} &= I(\gamma, \delta)^{-1} \int_0^\infty d\lambda \lambda^{-1+\delta} \sum_j |e_j(V) + \lambda|_-^\gamma. \end{aligned} \quad (9)$$

However, $e_j(V) + \lambda$ are the eigenvalues of the potential $V(x) + \lambda$. Therefore, by definition (3),

$$\sum_j |e_j(V) + \lambda|_-^\gamma \leq R_n(\gamma) \int d^n z |H(x, p) + \lambda|_-^\gamma \quad (10)$$

and, by substitution in (9) and using (8),

$$\begin{aligned} \sum_j |e_j(V)|_-^{\gamma+\delta} &\leq I(\gamma, \delta)^{-1} \int_0^\infty d\lambda \lambda^{-1+\delta} R_n(\gamma) \\ &\quad \times \int d^n z |H(x, p) + \lambda|_-^\gamma = R_n(\gamma) \int d^n z |H(x, p)|_-^{\gamma+\delta}. \end{aligned} \quad (11)$$

Hence

$$R_n(\gamma + \delta, V) \leq R_n(\gamma)$$

and therefore

$$R_n(\gamma + \delta) \leq R_n(\gamma). \quad (12)$$

(b) Let $\gamma > \max(0, 1 - n/2)$ and assume that for some V with $|V|_- \in L^{\gamma+n/2}(\mathbb{R}^n)$

$$R_n(\gamma, V) = R_n(\gamma). \quad (13)$$

In particular (13) implies that

$$0 > e_0 \equiv \inf \text{spec}(-\Delta + V) > \text{ess inf } \{V(x)\}. \quad (14)$$

We shall prove that $R_n(\cdot)$ is strictly decreasing from the left at γ by showing that

$$\begin{aligned} \liminf_{\delta \rightarrow 0^+} [R_n(\gamma - \delta) - R_n(\gamma)] / \delta \\ \geq R_n(\gamma) \frac{\int d^n z |H(x, p)|_-^\gamma \mu_\gamma(|e_0/H|)}{\int d^n z |H(x, p)|_-^\gamma} > 0, \end{aligned} \quad (15)$$

where

$$\mu_\gamma(t) = \int_t^1 d\lambda (1 - \lambda)^\gamma / \lambda > (1 - t)^{1+\gamma} / (1 + \gamma). \quad (16)$$

The key fact which will be used is that the integral in (9) can be cut off from above at $|e_0|$.

For any $\delta > 0$

$$\begin{aligned} \sum_j |e_j(V)|_-^\gamma &= I(\gamma - \delta, \delta)^{-1} \\ &\quad \times \int_0^{|e_0|} d\lambda \lambda^{-1+\delta} \sum_j |e_j(V) + \lambda|_-^{\gamma-\delta} \\ &\leq R_n(\gamma - \delta) I(\gamma - \delta, \delta)^{-1} \\ &\quad \times \int d^n z \int_0^{|e_0|} d\lambda \lambda^{-1+\delta} |H(x, p) + \lambda|_-^{\gamma-\delta} \\ &= R_n(\gamma - \delta) \int d^n z |H(x, p)|_-^\gamma \\ &\quad - R_n(\gamma - \delta) I(\gamma - \delta, \delta)^{-1} \\ &\quad \times \int d^n z \int_{|e_0|}^\infty d\lambda \lambda^{-1+\delta} |H(x, p) + \lambda|_-^{\gamma-\delta}. \end{aligned} \quad (17)$$

Therefore, using (2), (13) and $R_n(\gamma - \delta) \geq R_n(\gamma)$,

$$\begin{aligned}
& [R_n(\gamma - \delta) - R_n(\gamma)] / \delta \geq [R_n(\gamma) / \delta I(\gamma - \delta, \delta)] \\
& \times \int_{H < e_0} d^n z |H(x, p)|_-^\gamma \\
& \times \int_{|e_0/H|}^1 d\lambda \lambda^{-1+\delta} (1-\lambda)^{\gamma-\delta} \left/ \int d^n z |H(x, p)|_-^\gamma \right. . \quad (18)
\end{aligned}$$

Using $\lim_{\delta \rightarrow 0^+} \delta I(\gamma - \delta, \delta) = 1$ and Fatou's lemma, we obtain (15). ■

In view of (16) the theorem implies

Corollary 1: If for some γ , $R_n(\gamma) = 1$ then $R_n(\bar{\gamma})$

$= 1$ for all $\bar{\gamma} > \gamma$. Moreover, for $\bar{\gamma} > \gamma$ the supremum in (3) is not attained by any potential.

This proves part of a conjecture made in [2] (another part of the conjecture was disproved, for $n \geq 7$, in [7]).

In one dimension we can say even more since it is known, [2], that $R_1(3/2) = 1$ ($R_1(\gamma) > 1$ for $\gamma < 3/2$).

Corollary 2: For $\gamma \geq 3/2$, $R_1(\gamma) = 1$.

One may also study bounds like (3) for some re-

stricted classes of potentials, V , as was done in [7] for the spherically symmetric ones (the constants thus obtained are no larger than $R_n(\gamma)$ but it is not known whether any of them are strictly smaller). The theorem and its proof extend to such bounds as long as the class of potentials is closed under the addition of constants.

References

- [1] E.H. Lieb and W.E. Thirring, Phys. Rev. Lett. 35 (1975) 687. See Phys. Rev. Lett. 35 (1975) 1116 for errata. Also E.H. Lieb, Rev. Mod. Phys. 48 (1976) 553.
- [2] E.H. Lieb and W.E. Thirring, in: Studies in mathematical physics, Essays in honor of V. Bargmann (Princeton Univ. Press, Princeton, N.J., 1976).
- [3] A. Martin, Helv. Phys. Acta 45 (1972) 140.
- [4] H. Tamura, Proc. Japan Acad. 50 (1974) 19.
- [5] M. Cwikel, Ann. of Math. 106 (1977) 93.
- [6] E.H. Lieb, Bull. Amer. Math. Soc. 82 (1976) 751.
- [7] V. Glaser, H. Grosse and A. Martin, Bounds on the Number of Eigenvalues of the Schrödinger Operator, CERN preprint TH2432 (1977).
- [8] G.V. Rosenblum, The distribution of the discrete spectrum for singular differential operators, Izvestia Math. 164 No. 1 (1976) 75.
- [9] M.S. Birman and V.V. Borzov, On the asymptotics of the discrete spectrum of some singular differential operators, Topics in Math. Phys. 5 (1972) 19.

THE NUMBER OF BOUND STATES OF ONE-BODY
SCHROEDINGER OPERATORS AND THE WEYL PROBLEM

Elliott H. Lieb¹

ABSTRACT. If $\tilde{N}(\Omega, \lambda)$ is the number of eigenvalues of $-\Delta$ in a domain Ω in a suitable Riemannian manifold of dimension n , we derive bounds of the form $\tilde{N}(\Omega, \lambda) \leq D_n \lambda^{n/2} |\Omega|$ for all Ω, λ, n . Likewise, if $N_\alpha(V)$ is the number of nonpositive eigenvalues of $-\Delta + V(x)$ which are $\leq \alpha \leq 0$, then $N_\alpha(V) \leq L_n \int_M [V - \alpha]_-^{n/2}$ for all α and V and $n \geq 3$.

I. INTRODUCTION AND BACKGROUND.

Two closely related problems will concern us here : One is to bound the nonpositive eigenvalues of the one-body Schroedinger operator

$$H = -\Delta + V(x). \quad (1.1)$$

The other is to find an upper bound for

$$\tilde{N}(\Omega, \lambda) = \text{number of eigenvalues of } -\Delta \leq \lambda \quad (1.2)$$

in a domain Ω with Dirichlet boundary conditions.

In both cases the setting is a Riemannian manifold, M , and Δ is the Laplace-Beltrami operator. The only way in which the properties of the manifold will appear in our results will be through the fundamental solution of the heat equation

$$G(x, y; t) = [e^{x p(t \Delta)}](x, y) \quad (1.3)$$

1980 Mathematics Subject Classification 35P15.

¹ Work supported by U.S. National Foundation grants PHYS-7825390 and INT 78-01160.

evaluated on the diagonal $x = y$, or else through the Green function

$$(-\Delta + e)^{-1}(x, y) = \int_0^\infty dt e^{-et} G(x, y; t) \quad (1.4)$$

for $e \geq 0$. (Note: G is defined for the *whole* manifold, not the subdomain Ω in the case of (1.2). One could of course, use G defined for the domain Ω in all the formulas, but then the dependence of the result on Ω will be complicated. It is precisely to avoid this complication that we use the G for the whole of M).

Let us begin with the problem defined by (1.2), which we may term the *Weyl problem*. Weyl [1] proved the asymptotic formula (for suitable domains) :

$$\tilde{N}(\Omega, \lambda) \approx C_n \lambda^{n/2} |\Omega| \quad (1.5)$$

where $|\Omega|$ is the Riemannian volume of Ω , n is the dimension of M and

$$\begin{aligned} C_n &= (4\pi)^{-n/2} \Gamma(1 + n/2)^{-1} \\ &= (2\pi)^{-n} \tau_n, \end{aligned} \quad (1.6)$$

where τ_n is the volume of the unit ball in $\mathbb{R}^n$. The constant C_n is called the *classical constant* for reasons which will become clear later. This result (1.5) is discussed in [2] and [3], §XIII.15, vol.4. The proof uses Dirichlet-Neumann bracketing.

Polya's conjecture is that (1.5) holds in $\mathbb{R}^n$ for all λ and Ω , not just asymptotically. Here we will prove

THEOREM 1. *For all λ and Ω there exist constants D_n and E_n (depending on the manifold M) such that*

$$\tilde{N}(\Omega, \lambda) \leq D_n \lambda^{n/2} |\Omega| \quad (1.7)$$

$$\text{if} \quad G(x, x; t) \leq A_n t^{-n/2}, \quad \forall x \in M, \quad \forall t > 0 \quad (1.8)$$

for some $A_n < \infty$, while

$$\tilde{N}(\Omega, \lambda) \leq (D_n \lambda^{n/2} + E_n) |\Omega| \quad (1.9)$$

$$\text{if} \quad G(x, x; t) \leq A_n t^{-n/2} + B_n, \quad \forall x \in M, \quad \forall t > 0, \quad (1.10)$$

D_n and E_n are proportional to A_n and B_n respectively.

In particular (1.8) and (1.7) hold for $\mathbb{R}^n$ (with $A_n = (4\pi)^{-n/2}$) and for many noncompact M , e.g. homogeneous spaces with curvature ≤ 0 . (1.10) and (1.9) hold for compact M .

Next we turn to the Schroedinger problem (1.1). Let $E_1(V) \leq E_2(V) \leq \dots \leq 0$ be the nonpositive eigenvalues of H on $L^2(M)$. If we write

$$V = V_+ - V_- \quad (1.11)$$

$$\begin{aligned} \text{with} \quad V_+ &= |V(x)|, \quad \text{when } V(x) \geq 0 \\ &= 0, \quad \text{otherwise} \end{aligned} \quad (1.12)$$

BOUND STATES OF SCHROEDINGER OPERATORS AND THE WEYL PROBLEM 243

then the negative spectrum of H is discrete if $V_- \in L^{n/2}$, for example.

DEFINITION. $N_\alpha(V)$ is the number of eigenvalues of H which are $\leq \alpha \leq 0$.

Our main result is

THEOREM 2. Suppose (1.8) holds and suppose $n \geq 3$. Then

$$N_0(V) \leq L_n \int_M V_-(x)^{n/2} dx \quad (1.13)$$

for some constant L_n depending on M .

There are many remarks to be made about Theorem 2 and its connection with Theorem 1. First, the history of (1.13). Rosenbljum [4] first announced Theorem 2 in 1972. Unaware of this, Simon [5] proved an inequality of the form

$$N_0(V) \leq S_{n,\epsilon} [\|V_-\|_{n/2+\epsilon} + \|V_+\|_{n/2+\epsilon}]^{n/2}$$

with $S_{n,\epsilon} \rightarrow \infty$ as $\epsilon \downarrow 0$. Also unaware of [4], Cwikel and myself [6,7] simultaneously found a proof of Theorem 2. Reed and Simon [3] call Theorem 2 the Cwikel - Lieb - Rosenbljum bound. Cwikel's method exploits some ideas in [5]. All three methods are different, Cwikel's and Rosenbljum's are applicable to a wider class of operators than the Schroedinger operator, but my method [7], based on Wiener integrals, which is the one presented here, gives the best constant by far. This result was announced in [7] and the proof was written up in [3], Theorem XIII.12 and in [8]. Because all the technical details can easily be found in [8], the presentation given here will ignore technicalities. Not only am I indebted to B.Simon for his help with the technical details, as just mentioned, but I also wish to acknowledge his role in stimulating my interest in the problem, and for his constant encouragement while the ideas were taking shape.

The connection between the two theorems is

PROPOSITION 3. Let $\alpha \leq 0$. Then for all M

$$N(\Omega, \lambda) \leq N_\alpha((\alpha - \lambda) \chi_\Omega) \quad (1.14)$$

where χ_Ω is the characteristic function of Ω .

PROOF. Let ψ_j be the j th eigenfunction of $-\Delta$ in Ω and let $\tilde{\psi}_j$, defined on all of M , be ψ_j in Ω and zero outside. (1.14) is obtained by using the $\tilde{\psi}_j$ as variational functions for $-\Delta + (\alpha - \lambda) \chi_\Omega$ in the Rayleigh - Ritz variational principle. QED.

Similarly, we have

PROPOSITION 4. if $0 \leq \beta \leq -\alpha$, then for all M

$$N_\alpha(V) \leq N_{\alpha+\beta}(-[V + \beta]_-). \quad (1.15)$$

PROOF. Same as for proposition 3. Alternatively, one can remark that $V(x) \geq -[V + \beta]_-(x) - \beta$, and adding a positive operator can never decrease any eigenvalue. QED.

In particular, $N_\alpha(V) \leq N_0(-[V - \alpha]_-)$ and $N_0(V) \leq N_0(-V_-)$, whence we have

COROLLARY 5. If $n \geq 3$ and (1.13) holds, then

$$N_\alpha(V) \leq L_n \int_M [V - \alpha]_-(x)^{n/2} dx. \quad (1.16)$$

Moreover to prove (1.13) it is sufficient to consider V satisfying $V(x) \leq 0$, $\forall x$.

Proposition 3 will be used to derive Theorem 1 from Theorem 2 (actually from a generalization of Theorem 2, namely Theorem 8 and (4.3), which holds for all n). However, at this point we can, under the assumption that (1.8) holds, deduce (1.7) of Theorem 2 from proposition 3 and Theorem 2.

Choosing $\alpha = 0$, we have

$$\begin{aligned} N(\Omega, \lambda) &\leq L_n \lambda^{n/2} \int_M \chi_\Omega^{n/2} dx \\ &= L_n \lambda^{n/2} |\Omega| \end{aligned} \quad (1.17)$$

for $n \geq 3$.

It is to be emphasized that Theorem 2 is more delicate than Theorem 1. For one thing Theorem 1 holds for all n , whereas Theorem 2 holds only for $n \geq 3$. The analogue of Theorem 2 is definitely false for $n = 1$ and 2. In $\mathbb{R}^n$, at least, an arbitrarily small, nonpositive potential always has a negative eigenvalue when $n < 3$; cf. [3], Theorem XIII.11. For another thing, the best constant in (1.7) is, according to the Polya conjecture, the classical constant C_n (in $\mathbb{R}^n$ at least). However, the best constant L_n is bigger than C_n in general. It is easy to prove [9] (by considering a very "large" V and using Dirichlet-Neumann bracketing) that $L_n \geq C_n$. In [9] it was shown that $L_n > C_n$ for $3 \leq n \leq 7$ and recently [16] it was shown that $L_n > C_n$ for $n \geq 7$; in both cases explicit examples were constructed.

It is somewhat ironic that although Theorem 2 starts to hold only for $n = 3$, Theorem 1 is easiest to prove for $\mathbb{R}^n$ with $n = 1$. In that case the only domain that need be considered is a finite interval, and there the eigenvalues of $-\Delta$ can be computed explicitly. The Polya conjecture is easily seen to be true.

The intuition behind Theorem 2, and thereby the reason for calling C_n the classical constant is important. In the semiclassical picture of quantum mechanics in $\mathbb{R}^n$, which is similar to WKB theory, one has the mystical postulate that "each nice set in phase space $\Gamma = \{(p, x) \mid p \in \mathbb{R}^n, x \in \mathbb{R}^n\}$ of volume $(2\pi)^n$ can accommodate one eigenstate of H ". This postulate can be made more precise by mean of the Dirichlet-Neumann bracketing method mentioned before. In any event, since the "eigenvalues" of $-\Delta$ are p^2 and the "eigenvalues" of V are $V(x)$, the postulate implies that

BOUND STATES OF SCHROEDINGER OPERATORS AND THE WEYL PROBLEM 245

$$N_{\alpha}(V) \approx (2\pi)^{-n} \int_{\mathbb{R}^n \times \mathbb{R}^n} dp dx \, \Theta(\alpha - (p^2 + V(x))) \quad (1.18)$$

with $\Theta(a) = 1$, for $a \geq 0$ and $= 0$ for $a < 0$. The p integration in (1.18) for fixed x , is easy to do, namely

$$\int_{p^2 \leq \alpha - V(x)} dp = \tau_n [V - \alpha]_-(x)^{n/2} \quad (1.19)$$

Thus, (1.18) yields

$$N_{\alpha}(V) \approx c_n \int [V - \alpha]_-(x)^{n/2} dx \quad (1.20)$$

While the chief purpose of this paper is to prove Theorems 1 and 2, quantities of no less interest are the moments of the nonpositive eigenvalues of H .

DEFINITION. For $\gamma > 0$

$$\begin{aligned} I_{\gamma}(V) &= \sum_i |E_i(V)|^{\gamma} \\ &= \gamma \int_0^{\infty} |\alpha|^{\gamma-1} N_{\alpha} d\alpha \end{aligned} \quad (1.21)$$

$I_0(V)$ is defined to be $N_0(V)$.

For $n \geq 3$ we can use Corollary 5 and Fubini's theorem to obtain

$$\begin{aligned} I_{\gamma}(V) &\leq \gamma L_n \int_M dx \int_{-\infty}^0 |\alpha|^{\gamma-1} [V - \alpha]_-(x)^{n/2} d\alpha \\ &= \gamma L_n \int_M dx \int_{-V_-(x)}^0 d\alpha |\alpha|^{\gamma-1} (V_-(x) + \alpha)^{n/2} \\ &= L_{\gamma,n} \int_M V_-(x)^{\gamma+n/2} dx \end{aligned} \quad (1.22)$$

with

$$L_{\gamma,n} = L_n \Gamma(\gamma+1) \Gamma(1+n/2) \Gamma(1+\gamma+n/2)^{-1} \quad (1.23)$$

There are several things to be said about (1.22). Although it was derived from Corollary 5 under the assumptions (1.13) and $n \geq 3$, it holds much more generally. For example it holds in $\mathbb{R}^n$ for $n = 2$, $\gamma > 0$ and for $n = 1$, $\gamma > 1/2$ provided $\alpha > 0$. This was first given in [9]. In [9] it was stated that it holds for $n = 1$ and $\gamma = 1/2$. That was an error; it is not known if (1.22) holds for $n = 1$, $\gamma = 1/2$ but it is known [9] that (1.22) does not hold for $n = 1$, $\gamma < 1/2$. In section II we shall briefly mention how to deduce (1.22).

The best constant $L_{\gamma,n}$ in (1.22) is not given by (1.23), as the foregoing remark already indicates. If we use C_n in place of $L_{\gamma,n}$ in (1.23) we have the classical value of $L_{\gamma,n}$ namely,

$$L_{\gamma,n}^C = (4\pi)^{-n/2} \Gamma(\gamma+1) \Gamma(1+\gamma+n/2)^{-1} \quad (1.24)$$

As in the case of L_n , it is easy to prove that $L_{\gamma,n} \geq L_{\gamma,n}^C$. The classical constant $L_{\gamma,n}^C$ can also be "derived" from the semiclassical assumption as in (1.18), (1.19), namely

$$\sum_1 |E_1|^\gamma \approx (2\pi)^{-n} \int_{\mathbb{R}^n} |p^2 + V(x)|^\gamma \Theta(-p^2 - V(x)) \, dp \, dx. \quad (1.25)$$

If the p integration is done in (1.25), the result is (1.22) with $L_{\gamma,n}^C$.

An important question is :

$$\text{When is } L_{\gamma,n} = L_{\gamma,n}^C ?$$

It seems to be true that $L_{\gamma,n} = L_{\gamma,n}^C$ for γ large enough, depending on n . This is known to be true [9, 10] for $n = 1$ and $\gamma \geq 3/2$. In fact [10], if $L_{\gamma_0,n} = L_{\gamma_0,n}^C$ for some γ_0 , then equality holds for all $\gamma \geq \gamma_0$.

The case of primary *physical interest* is $\gamma = 1$, $n = 3$, where it is conjectured [9] that equality holds. If this were so, it would have important consequences for physics and it is hoped that someone will be motivated to solve the problem.

We now turn to the proof of Theorems 1 and 2 in the next three sections.

II. THE BIRMAN-SCHWINGER KERNEL

As stated in Corollary 5 we can assume $V(x) \leq 0$. Therefore write $V(x) = -U(x)$, $U \geq 0$.

A useful device for studying the nonpositive eigenvalues of $-\Delta - U$ was discovered by Birman [11] and Schwinger [12]. If $(-\Delta - U)\psi = E\psi$, $E \leq 0$, then

$$\psi = (-\Delta + |E|)^{-1} U \psi. \quad (2.1)$$

Defining $U^{1/2}\psi = \phi$, and multiplying (2.1) by $U^{1/2}$, we have

$$\phi = K_{|E|}(U)\phi \quad (2.2)$$

where $K_e(U)$, for $e \geq 0$, is the positive *Birman-Schwinger Kernel* given explicitly by

$$K_e(x, y; U) = U(x)^{1/2} (-\Delta + e)^{-1}(x, y) U^{1/2}(y). \quad (2.3)$$

What (2.2) says is that for every nonpositive eigenvalue, E , of $-\Delta - U$, $K_{|E|}(U)$ has an eigenvalue 1. The converse is also easily seen to hold (see [3,8] for more details). $K_e(U)$ is to be thought of as an operator on L^2 ; we will see that it is compact, when $e > 0$ at least, and U is in a suitable L^p space.

In addition to the advantage that the study of the E 's reduces to the study of a compact operator, there is the following important fact: Since $(-\Delta + e)^{-1}$ is operator monotone decreasing as a function of e , so is K_e . Hence (with $V = -U$),

$$N_\alpha(V) = k_{|\alpha|}(U) \equiv \text{number of eigenvalues of } K_{|\alpha|}(U) \geq 1. \quad (2.4)$$

(2.4) will be exploited in the following way.

PROPOSITION 6. Let $F: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be any function such that $F(x) \geq 1$ for $x \geq 1$.

BOUND STATES OF SCHROEDINGER OPERATORS AND THE WEYL PROBLEM 247

Then

$$\begin{aligned} \text{Tr } F(K_e(U)) &= \sum_i F(l_e^i(U)) \\ &\geq k_e(U) = N_{-e}(V) \end{aligned} \quad (2.5)$$

where Tr means trace and the $l_e^i(U)$ are the eigenvalues of $K_e(U)$.

For example, consider $\mathbb{R}^n$, $n \leq 3$, and $F(x) = x^2$. Then

$$\begin{aligned} \text{Tr } K_e(U)^2 &= \iint U(x) U(y) [(-\Delta + e)^{-1}(x - y)]^2 dx dy \\ &\leq \|U\|_2^2 \|(-\Delta + e)^{-1}(x)\|_2^2 \end{aligned} \quad (2.6)$$

by Young's inequality. The last factor is of the form

$$\|(-\Delta + e)^{-1}(x)\|_2^2 = h_n e^{-2+n/2} \quad (2.7)$$

This shows that $K_e(U)$ is Hilbert-Schmidt when $U \in L^2$ and $e > 0$. When $n > 3$, $h_n = \infty$ but one can show that $K_e(U)$ is compact by considering the trace of a higher power of $K_e(U)$, cf. [9].

At this point we can derive the aforementioned bound, (1.22), on $I_Y(V)$. If we use (2.5) and (2.6) and insert the latter in (1.21), the α integration will diverge. The trick [9] is to use Proposition 4 with $2\beta = e = -\alpha$. Thus $N_{-e}(V) \leq \text{Tr } K_{e/2}([V + e/2]_-)$

$$\leq \| [V - e/2]_+ \|_2^2 h_n (e/2)^{-2+n/2} \quad (2.8)$$

Inserting (2.8) into (1.21), and doing the α integration first, we obtain

$$\begin{aligned} I_Y(V) &\leq 2^{2-n/2} \gamma h_n \int dx \int_{-2U(x)}^0 d\alpha |\alpha|^{\gamma-3+n/2} [U(x) + \alpha/2]^2 \\ &= L_{\gamma,n} \int U(x)^{\gamma+n/2} dx \end{aligned} \quad (2.9)$$

if $\gamma > 2 - n/2$ and $n \leq 3$.

(2.9) can be extended to other values of n and γ (but with $\gamma > 1/2$ for $n = 1$), cf. [9].

There is no way, however, to make this method work with $F(x) = x^a$ when $\gamma = 0$ and $n \geq 3$. Quiet reflection shows that if theorem 2 is to be provable by this method then we need $x^{-n/2} F(x) \rightarrow 0$ as $x \rightarrow 0$ but $x^{-1} F(x)$ remains bounded as $x \rightarrow \infty$. The tool we will use to bound $\text{Tr} F(K_e(U))$ for such F 's is the Wiener integral. That is the subject of the next section, which is really the main point of this paper.

III. THE WIENER INTEGRAL REPRESENTATION OF $F(K_e(U))$

Let $d\mu_{x,y;t}$ be conditional Wiener measure on paths $\omega(t)$ with $\omega(0) = x$ and $\omega(t) = y$. This measure gives a representation for G (cf. (3)) by

$$\int d\mu_{x,y;t}(\omega) = G(x,y;t) = e^{t\Delta}(x,y). \quad (3.1)$$

G itself has the *semigroup property*

$$\int_M G(x, y; t) G(y, z; s) dy = G(x, z; t+s). \quad (3.2)$$

The well known Feynman-Kac formula [13] is

$$\int d\mu_{x,y;t}(\omega) \exp[-\lambda \int_0^t U(\omega(s)) ds] = e^{-t(-\Delta + \lambda U)}(x, y). \quad (3.3)$$

(Note the signs in (3.3).). Take $e \geq 0$, $\lambda \geq 0$, and $U \geq 0$, multiply (3.3) by $U(x)^{1/2} U(y)^{1/2} \exp(-et)$ and integrate with respect to t . The result is

$$\begin{aligned} A &= U(x)^{1/2} U(y)^{1/2} \int_0^\infty dt e^{-et} \int d\mu_{x,y;t}(\omega) \exp[-\lambda \int_0^t U(\omega(s)) ds] \\ &= \{U^{1/2}(-\Delta + \lambda U + e)^{-1} U^{1/2}\}(x, y). \end{aligned} \quad (3.4)$$

Now

$$(-\Delta + e)^{-1} = (-\Delta + \lambda U + e)^{-1} + \lambda(-\Delta + \lambda U + e)^{-1} U(-\Delta + e)^{-1} \quad (3.5)$$

Multiplying (3.5) on both sides by $U^{1/2}$ we obtain

$$A = \{K_e(U) [1 + \lambda K_e(U)]^{-1}\}(x, y). \quad (3.6)$$

(3.6) can be cast in a more general form. If $g(x) = \exp(-\lambda x)$ and $F(x) = x(1 + \lambda x)^{-1}$ then

$$F(x) = x \int_0^\infty dy e^{-y} g(xy) \quad (3.7)$$

and

$$\begin{aligned} U(x)^{1/2} U(y)^{1/2} \int_0^\infty dt e^{-et} \int d\mu_{x,y;t}(\omega) g(\int_0^t U(\omega(s)) ds) \\ = F(K_e(U))(x, y). \end{aligned} \quad (3.8)$$

Next we want to take the trace of both sides of (3.8). This is obtained by setting $y = x$ and integrating. (To be precise, one must first take $U \in C^\infty$ and then use a limiting argument.) The point to notice is that the x dependence occurs through

$$\int_M dx U(x) d\mu_{x,x;t}(\omega) = \int_M dx d\mu_{x,x;t}(\omega) U(\omega(0)). \quad (3.9)$$

By the semigroup property of $\exp[-t(-\Delta + \lambda U)]$, however, the same result is obtained (after the $d\mu$ integration) if $U(\omega(0))$ in (3.9) is replaced by $U(\omega(s))$ for any $0 \leq s \leq t$. Thus,

$$\int_0^\infty dt t^{-1} e^{-et} \int_M dx \int d\mu_{x,x;t}(\omega) f(\int_0^t U(\omega(s)) ds) = \text{Tr } F(K_e(U)). \quad (3.10)$$

with

$$f(x) = xg(x), \quad (3.11)$$

i.e.

$$F(x) = \int_0^\infty dy y^{-1} e^{-y} f(xy) \quad (3.12)$$

Now the relations between F , f and g are linear, as are the relations (3.8) and (3.10). The latter therefore extend to a large class of f 's. Since we are here not particularly interested in the operator version, (3.8), we will

BOUND STATES OF SCHROEDINGER OPERATORS AND THE WEYL PROBLEM 249

concentrate on the trace version (3.10). The g 's of the form $\exp(-\lambda x)$ are norm dense in the continuous functions which vanish at infinity. Furthermore, using the semigroup property of $\exp[-t(-\Delta + U)]$, one can see explicitly that (3.8) and (3.10) hold for f 's of the form $x^k \exp(-\lambda x)$, with k a positive integer. By a monotone convergence argument one arrives at

THEOREM 7. *Let f be a nonnegative lower semicontinuous function on $[0, \infty)$ satisfying:*

- (i) $f(0) = 0$
- (ii) $x^p f(x) \rightarrow 0$ as $x \rightarrow \infty$ for some $p < \infty$.

Let $U \geq 0$ and $U \in L^p + L^q$ with $p = n/2$ ($n \geq 3$), $p > 1$ ($n = 2$), $p = 1$ ($n = 1$) and $p < q < \infty$. Then (3.10) holds, with F given by (3.12), in the sense that both sides may be $+\infty$

The reader is referred to [8] for details. Obviously the class of f 's in Theorem 7 is not the largest possible, but it is more than adequate for our intended application.

The remark that allows us to bound the left side of (3.10) is the following.

THEOREM 8. *Suppose that f satisfies the conditions of Theorem 7 and that f is also convex. Then*

$$\text{Tr } F(K_e(U)) \leq \int_0^\infty dt \, t^{-1} e^{-et} \int_M dx \, G(x, x; t) f(tU(x)). \quad (3.13)$$

PROOF. By Jensen's inequality, for any fixed path $t \rightarrow \omega(t)$,
 $f(\int_0^t U(\omega(s)) ds) = f(\int_0^t (t^{-1} ds) t U(\omega(s))) \leq \int_0^t (t^{-1} ds) f(tU(\omega(s)))$.
 By the same remark as that preceding (3.10), $\int_M dx \int d\mu_{x,x;t}(\omega) h(\omega(s))$ is independent of s for any fixed function h . Inserting this in (3.10) gives (3.11). QED.

IV. APPLICATIONS OF THEOREM 8

PROOF OF THEOREM 1: We use Proposition 3 and Proposition 6. f is chosen to be of the following form for some $0 < a < b$

$$\begin{aligned} f(x) &= 0, & 0 \leq x \leq a \\ &= b(x - a), & a \leq x. \end{aligned} \quad (4.1)$$

F given by (3.12) is monotone, and the condition that $F(x) = 1$ is that a and b are related by

$$1 = h(a, b) = b \int_a^\infty dy (1 - a/y) e^{-y} = be^{-a} - abE_1(a) \quad (4.2)$$

where E_1 is the exponential integral. Assuming (1.10), (1.9) will be proved; (1.7) follows from the special assumption $B_n = 0$ in what follows.

In Proposition 3 write $\alpha = -e$ so that $\hat{N}(\Omega, \lambda) < (3.13)$ with $U = (e + \lambda)\chi_\Omega$. The x integration can be done last in (3.13) in which case $f(tU(x)) = 0$ if $x \notin \Omega$. For $x \in \Omega$ we change variables to $t \rightarrow t/(e + \lambda)$. Thus,

$$b|\Omega| \int_a^\infty dt (1 - a/t) (A_n(e + \lambda)^{n/2} t^{-n/2} + B_n) \exp[-et/(e + \lambda)] \geq \hat{N}(\Omega, \lambda). \quad (4.3)$$

The simplest choice for e , which is arbitrary, is $e = 0$. This will work only when $n \geq 3$ and $B_n = 0$. But for all cases we can choose $e = c\lambda$, $c > 0$ and thereby prove Theorem 1. QED.

PROOF OF THEOREM 2: Proposition 6 is used again and the proof parallels that of Theorem 1. (4.1) and (4.2) are assumed. We change variables in (3.13) to $t \rightarrow tU(x)^{-1}$ if $U(x) \neq 0$. If $e = 0$ the result is (1.13) with

$$L_n = A_n \int_0^\infty dt t^{-1-n/2} f(t) = 4A_n b a^{1-n/2} [n(n-2)]^{-1}. \quad \text{QED.} \quad (4.4)$$

REMARKS: (i) If $\alpha = -e \neq 0$ then the only estimate we have for $N_\alpha(V)$ is contained in Corollary 5, which is valid only for $n \geq 3$. Alternatively, one could try to estimate (3.13) directly with $e \neq 0$, but this is messy. As stated earlier, no inequality of the form (1.16) holds for all α, V when $n = 1$ or 2 . But recently Itô [14] has bounded (3.13) when $e \neq 0$ and $n = 2$. He uses the fact that $f(x) \leq bx$ for $x \geq a$ and obtains a complicated upper bound for $N_\alpha(V)$ in terms of $\|V_-\|_2$ and $\|V_-\ln V_-\|_2^{1/2}$.

(ii) If $n \geq 3$ and $B_n = 0$ we can choose $e = 0$ in (4.3).

This estimate for D_n is, of course, the same as L_n given by (4.4).

As an illustration of how good our bound is let us consider the case of $\mathbb{R}^3$, where $A_3 = (4\pi)^{-3/2}$. We choose $a = 0.25$ in (4.1) and find that $E_1(.25) = 1.0443$ and $b = 1.9315$ according to (4.2). Using (4.4),

$$D_3 = L_3 = 0.1156. \quad (4.5)$$

This value of D_3 can be used in (1.7). When compared with $C_3 = 0.0169$, which is supposed to be the sharp constant, it is not very good. The estimate for D_3 can be improved by using (4.3) with $e = c\lambda$, $c > 0$.

If, however, the same number, L_3 , is used in (1.13) the result is quite good. As already stated, the best $L_3 > C_3$. In fact, by an explicit example,

$$L_3 \geq (3\pi)^{-3/2} \Gamma(3)/\Gamma(3/2) = 0.0780, \quad (4.6)$$

BOUND STATES OF SCHROEDINGER OPERATORS AND THE WEYL PROBLEM 251

cf. [9], eqn. (4.24). It is conjectured that the right side of (4.6) is, in fact, the sharp constant in (1.13) for $\mathbb{R}^3$. In any case our result, (4.5), is off by at most 49%.

As stated in Section 1, a quantity of physical interest is I_1 , the sum of the absolute values of the eigenvalues, in $\mathbb{R}^3$. Using the bound (1.22), (1.23), together with (4.5), we have

$$L_{1,3} \leq (2/5)L_3 = .04624 \quad (4.7)$$

This result was announced in [15].

BIBLIOGRAPHY

1. H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte Linearer partieller Differentialgleichungen", Math. Ann. 71 (1911), 441-469.
2. M. Kac, "Can one hear the shape of a drum?", Slaughter Memorial Papers, no. 11, Amer. Math. Monthly 73 (1966), no. 4, part II, 1-23.
3. M. Reed and B. Simon, Methods of Modern Mathematical Physics, Acad. Press, N. Y., 1978.
4. G. V. Rosenbljum, "Distribution of the discrete spectrum of singular differential operators", Dokl. Akad. Nauk SSSR, 202 (1972), 1012-1015 (MR 45 #4216). The details are given in "Distribution of the discrete spectrum of singular differential operators", Izv. Vyss. Uchebn. Zaved. Matematika 164 (1976), 75-86. [English trans. Sov. Math. (Iz. VUZ) 20 (1976), 63-71.]
5. B. Simon, "Weak trace ideals and the number of bound states of Schroedinger operators", Trans. Amer. Math. Soc. 224 (1976), 367-380.
6. M. Cwikel, "Weak type estimates for singular values and the number of bound states of Schroedinger operators", Ann. Math. 106 (1977), 93-100.
7. E. Lieb, "Bounds on the eigenvalues of the Laplace and Schroedinger operators", Bull. Amer. Math. Soc. 82 (1976), 751-753.
8. B. Simon, Functional Integration and Quantum Physics, Academic Press, N. Y., to appear 1979.
9. E. Lieb and W. Thirring, "Inequalities for the moments of the eigenvalues of the Schroedinger equation and their relation to Sobolev inequalities", in Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann (E. Lieb, B. Simon and A. Wightman eds.), Princeton Univ. Press, Princeton, N. J., 1976. These ideas were first announced in "Bound for the kinetic energy of fermions which proves the stability of matter", Phys. Rev. Lett. 35 (1975), 687-689, Errata 35 (1975), 1116.
10. M. Aizenman and E. Lieb, "On semi-classical bounds for eigenvalues of Schroedinger operators", Phys. Lett. 66A (1978), 427-429.
11. M. Birman, "The spectrum of singular boundary problems", Math. Sb. 55 (1961), 124-174. (Amer. Math. Soc. Trans. 53 (1966), 23-80).
12. J. Schwinger, "On the bound states of a given potential", Proc. Nat. Acad. Sci. U.S.A. 47 (1961), 122-129.

13. M. Kac, "On some connections between probability theory and differential and integral equations", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Univ. of Calif. Press, Berkeley, 1951, 189-215.

14. K. R. Ito, "Estimation of the functional determinants in quantum field theories", Res. Inst. for Math. Sci., Kyoto Univ. (1979), preprint.

15. E. Lieb, "The stability of matter", Rev. Mod. Phys. 48 (1976), 553-569.

16. V. Glaser, H. Grosse and A. Martin, "Bounds on the number of eigenvalues of the Schroedinger operator", Commun. Math. Phys. 59 (1978), 197-212.

DEPARTMENTS OF MATHEMATICS AND PHYSICS
PRINCETON UNIVERSITY
JADWIN HALL
P.O. BOX 708
PRINCETON, N. J. 08544

PHYSICAL REVIEW LETTERS

VOLUME 46

16 FEBRUARY 1981

NUMBER 7

Variational Principle for Many-Fermion Systems

Elliott H. Lieb

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08544

(Received 10 December 1980)

If ψ is a determinantal variational trial function for the N -fermion Hamiltonian, H , with one- and two-body terms, then $e_0 \leq \langle \psi, H\psi \rangle = E(K)$, where e_0 is the ground-state energy, K is the one-body reduced density matrix of ψ , and $E(K)$ is the well-known expression in terms of direct and exchange energies. If an arbitrary one-body K is given, which does not come from a determinantal ψ , then $E(K) \geq e_0$ does not necessarily hold. It is shown, however, that if the two-body part of H is positive, then in fact $e_0 \leq e_{\text{HF}} \leq E(K)$, where e_{HF} is the Hartree-Fock ground-state energy.

PACS numbers: 05.30.Fk, 21.60.Jz, 31.15.+q

The variational principle is useful for obtaining accurate upper bounds to the ground-state energy e_0 of an N -particle fermion Hamiltonian, H_N . A normalized wave function ψ_N (or density matrix ρ_N) which satisfies the Pauli principle is required; then $e_0 \leq e(\rho_N) \equiv \text{Tr} \rho_N H_N$ (with $\rho_N = |\psi_N\rangle \langle \psi_N|$ being a pure state in the wave-function case). In practice, however, it is often possible to make a good guess for ρ_N^{-1} , the reduced single-particle density matrix, but the evaluation of $e(\rho_N)$ is complicated by the reconstruction problem for ρ_N : To evaluate $e(\rho_N)$ we first have to know ρ_N . In the simplest case ρ_N^{-1} is an N -dimensional projection and ρ_N is a pure state, with ψ_N being a determinantal, or Hartree-Fock (HF) function. Otherwise, ρ_N is a very complicated (and, in general, a nonunique) function of ρ_N^{-1} , and the calculation of $e(\rho_N)$ can be extremely difficult because of the "orthogonality problem." For this reason most variational calculations do not depart very far from a HF calculation.

It is the purpose of this note to show that under a positivity condition on the two-body part of H_N (which, fortunately, holds for one case of major interest—the Coulomb potential) it is possible to obtain an upper bound to e_0 which involves only ρ_N^{-1} ; the reconstruction problem is eliminated. In the HF case, this bound agrees with $e(\rho_N)$. Moreover, our bound, E , satisfies $E \geq e_{\text{HF}}$,

where e_{HF} is the lowest HF energy. While our bound is thus not superior to the best HF bound, it may be superior in practice because the exact HF orbitals are unknown in general. A possible application might occur in the theory of itinerant ferromagnetism.

Let us make some definitions. Let $z = (x, \sigma)$ denote a single-particle space-spin variable and $\int dz \equiv \sum_{\sigma} \int dx$. A single-particle operator $K(z; z')$ is called admissible if it is positive semidefinite and

$$\text{Tr} K = N, \quad K \leq I, \quad \text{i.e.,}$$

$$\langle \psi, K\psi \rangle \leq \langle \psi, \psi \rangle \quad \text{for all } \psi. \quad (1)$$

Given ρ_N satisfying the Pauli principle,

$$\begin{aligned} \rho_N^{-1}(z; z') &\equiv N \int \rho_N(z, z_2, \dots, z_N; z', z_2, \dots, z_N) \\ &\quad \times dz_2 \dots dz_N. \end{aligned} \quad (2)$$

Any such ρ_N^{-1} is admissible. Conversely, given an admissible K there is always at least one ρ_N with $\rho_N^{-1} = K$. In the HF case $\rho_N = |\psi_N\rangle \langle \psi_N|$ and

$$\begin{aligned} \psi_N &= (N!)^{-1/2} \det[f_i(z_j)], \\ \rho_N^{-1}(z; z') &= \sum_{j=1}^N f_j(z) f_j^*(z'), \end{aligned} \quad (3)$$

with $f_1, \dots, f_N$ being any N orthonormal functions.

Consider now Hamiltonians of the form

$$H_N = \sum_{j=1}^N h_j + \sum_{1 \leq i < j \leq N} v_{ij}, \quad (4)$$

where h and v are self-adjoint operators. v is the two-body part and h is the one-body part [usually $-(\hbar^2/2m)\Delta + U(z)$]. Our method has obvious extensions to higher than two-body interactions, but for brevity only (4) will be considered. Normally v is diagonal [a local potential, such that $v_{ij} = v(z_i, z_j)$], but this is not necessary. If ρ_N is of the HF type, then ρ_N^2 , defined analogously to (2), satisfies $\rho_N^2 = K_2$, where

$$K_2(z, w; z', w') \\ \equiv K(z; z')K(w; w') - K(z; w')K(w; z'), \quad (5)$$

with $K = \rho_N^{-1}$. In this HF case $e(\rho_N) = E(K)$, where

$$E(K) \equiv \text{Tr}(K\hbar) + \frac{1}{2} \text{Tr}(K_2 v) \quad (6)$$

and

$$\text{Tr}(K_2 v) = \iint v(z, w) K_2(z, w; z, w) dz dw \quad (7)$$

in the diagonal case. This formula is well known. For any admissible K , (5) and (6) define $E(K)$.

The problem addressed here is the following: Given an arbitrary, admissible K , does a ρ_N exist such that $\rho_N^{-1} = K$ and $e(\rho_N) \leq E(K)$? If $v \geq 0$, the answer is yes! Note, however, that $\text{Tr} \rho_N^2 = N(N-1)$ but that $\text{Tr} K_2 = (\text{Tr} K)^2 - \text{Tr} K^2$, and this is $N(N-1)$ if and only if K is an N -dimensional projection as in (3). Otherwise, $\text{Tr} K_2 > N(N-1)$. Therefore the ρ_N which I wish to construct cannot simply satisfy $\rho_N^2 = K_2$. The solution must be more complicated than that.

Our main result is stated as follows:

Theorem.—Let v be positive semidefinite [i.e., in the diagonal case, $v(z, w) \geq 0$ for all z, w], and let K be any admissible single-particle operator. Then (i) there exists a density matrix ρ_N satisfying the Pauli principle such that $\rho_N^{-1} = K$ and

$$e_0 \leq e(\rho_N) \leq E(K); \quad (8)$$

(ii) there exists a normalized determinantal function ψ_N such that

$$e_0 \leq \langle \psi_N, H_N \psi_N \rangle \leq e(\rho_N) \leq E(K). \quad (9)$$

Note that in (ii) it is not claimed that if $\tilde{\rho}_N \equiv |\psi_N\rangle\langle\psi_N|$ then $\tilde{\rho}_N^{-1} = K$. However, (9) does say that, among all admissible K , an HF-type K [N-

dimensional projection (3)] gives the lowest value of $E(K)$. The proof requires the following:

Lemma.—Let $c_1 \geq c_2 \geq \dots$ be an infinite sequence with $0 \leq c_i \leq 1$ and $\sum_{i=1}^{\infty} c_i = N$, where N is an integer. Then there exist N orthonormal vectors $V^1, \dots, V^N$ in l^2 , the Hilbert space of square-summable sequences indexed by the positive integers (i.e., $V \in l^2$ means $\sum_{i=1}^{\infty} |V_i|^2 < \infty$), such that $\sum_{i=1}^N |V_j^i|^2 = c_j$.

Proof: Induction on N is used. For $N=1$, choose $V_j^1 = c_j^{1/2}$. Assume that the lemma holds for $N=n-1$; the lemma will first be proved for n under the assumption that $c_j = 0$ for $j \geq 2n$. Define $d_j \equiv 1 - c_j$ ($1 \leq j < 2n$), and $d_j = 0$ ($j \geq 2n$). The d_j satisfy the hypothesis for $n-1$, so that there exist orthonormal $W^1, \dots, W^n$ with $\sum_{i=1}^{n-1} |W_j^i|^2 = d_j$. Let $v^1, \dots, v^n$ be n orthonormal, $(2n-1)$ -dimensional vectors which are orthogonal to $W^1, \dots, W^{n-1}$ [thought of as $(2n-1)$ -dimensional vectors]. Then the n vectors $V_j^i = v_j^i$ ($1 \leq j < 2n$) [$V_j^i = 0$ ($j \geq 2n$)] satisfy the lemma. Next, suppose $c_j = 0$ for $j \geq J$. I use induction on J starting with $J=2n$. Note that $c_{J-1} + c_J \leq 1$ when $J \geq 2n$. For $J+1$, apply the lemma (with n and J) to the sequence $b_j = c_j$ ($1 \leq j < J-1$), $b_{J-1} = c_{J-1} + c_J$, and $b_j = 0$ ($j \geq J$). This sequence may not be decreasing, but that is irrelevant. Let $W^1, \dots, W^n$ be the orthonormal vectors. The required vectors, V^i , for $J+1$, are given by $V_j^i = W_j^i$ ($1 \leq j < J-1$), $V_{J-1}^i = W_{J-1}^i (c_{J-1}/b_{J-1})^{1/2}$, $V_J^i = W_{J-1}^i (c_J/b_{J-1})^{1/2}$, $V_j^i = 0$ ($j \geq J$). Finally, if $c_j > 0$ for all j , choose L so that $b_L \equiv \sum_{j=L}^{\infty} c_j \leq 1$. Then apply the lemma to the finite sequence of length L : $b_j = c_j$ ($1 \leq j < L$), b_L . If $W^1, \dots, W^n$ are the orthonormal vectors, let $V_j^i = W_j^i$ ($1 \leq j < L$), $V_j^i = W_L^i (c_j/b_L)^{1/2}$ for $j \geq L$. Q.E.D.

Proof of Theorem: Write $K(z; z') = \sum_{j=1}^{\infty} c_j f_j(z) \times f_j(z')^*$, where the f_j are the orthonormal eigenfunctions of K ("natural orbitals") and the eigenvalues of K , c_j , satisfy the hypothesis of the lemma. Let $V^1, \dots, V^N$ be the vectors of the lemma and let $\theta = \{\theta_1, \theta_2, \dots\}$ be any infinite sequence of reals. The N functions $F_k^{\theta}(z) = \sum_{j=1}^N e^{i\theta_j} V_j^k f_j(z)$ are orthonormal for any θ . Let $\rho_N^{\theta} = |\psi_N^{\theta}\rangle\langle\psi_N^{\theta}|$, where $\psi_N^{\theta}(z_1, \dots, z_N) = (N!)^{-1/2} \det[F_k^{\theta}(z_j)]$. Let $\langle \dots \rangle_{\theta}$ denote the average [over $[0, 2\pi]^{2N}$] with respect to all the θ_j . (Formally, this requires infinitely many integrations $\int_0^{2\pi} d\theta_j/2\pi$, but one can easily make sense of this by taking suitable limits.) It is easy to check, with use of the property of the V^i , that $\rho_N = \langle \rho_N^{\theta} \rangle_{\theta}$ satisfies $\rho_N^{-1} = \langle \rho_N^{\theta^{-1}} \rangle_{\theta} = K$. Now $\rho_N^2 \neq K_2$, as stated before, but $\rho_N^2 = \langle \rho_N^{\theta^{-2}} \rangle_{\theta} = K_2 - L_2$, with

$$2L_2(z, w; z', w') = \sum_{a,b=1}^{\infty} W_{ab} [f_a(z) f_b(w) - f_b(z) f_a(w)] [f_a^*(z') f_b^*(w') - f_b^*(z') f_a^*(w')]$$

and $W_{ab} = |\sum_{i=1}^N V_a^i V_b^{i*}|^2 \geq 0$. Thus, $e(\rho_N) = E(K) - D$, with $2D = \text{Tr}(L_2 v)$. But clearly L_2 is positive semidefinite, so that $D \geq 0$. This proves (i). To prove (ii), note that $E(K) \geq e(\rho_N) = \langle G^\theta \rangle_\theta$, where $G^\theta = \langle \psi_N^\theta, H_N \psi_N^\theta \rangle$ is real for each θ . Hence, for some θ , $G^\theta \leq e(\rho_N)$. Q.E.D.

A very useful discussion with Professor J. K. Percus is gratefully acknowledged. This work was partially supported by the National Science Foundation under Grant No. PHY-78-25390-A01.

Note added.—After reading this manuscript, Professor M. B. Ruskai kindly pointed out that the lemma is essentially a consequence of Horn's theorem¹: Let $y_1 \geq y_2 \geq \dots \geq y_M$ and $x_1 \geq x_2 \geq \dots \geq x_M$ be two sets of reals. Then there exists an

$M \times M$ hermitean matrix B with eigenvalues $\{x_i\}$ and diagonal elements $B_{ii} = y_i$, if and only if $\sum_{i=1}^t (x_i - y_i) \geq 0$ for all $1 \leq t \leq M$, and with equality for $t = M$. The existence of B is equivalent to $y_j = \sum_{i=1}^M |U_{ij}|^2 x_i$ for some unitary U . To apply this to the lemma, suppose that $c_j = 0$ for $j > M \geq N$ and take $y_j = c_j$ (for $j \leq M$) and $x_1 = x_2 = \dots = x_N = 1$, and $x_j = 0$ for $j > N$. The required orthonormal vectors V^i are then $V_j^i = U_{ij}$ for $j \leq M$ and $V_j^i = 0$ for $j > M$. Finally, if $c_j > 0$ for all j , then an argument such as that given at the end of the proof of the lemma, or something similar, must be used.

¹A. Horn, Am. J. Math. **76**, 620 (1954).

This paper is most properly cited as

Elliott H. Lieb, Phys. Rev. Lett. **46**, 457 (1981), and **47**, 69(E) (1981).

All corrections in the Erratum have been incorporated in this version of the reprints.

ERRATUM

VARIATIONAL PRINCIPLE FOR MANY-FERMI-
ON SYSTEMS. Elliott H. Lieb [Phys. Rev. Lett.
46, 457 (1981)].

The name of the author was misspelled in the
printed version. The correct spelling is as given
above.

Since the meaning of the abstract was adversely
affected by editorial processing, the entire ab-
stract is reproduced below for clarification:

If ψ is a determinantal variational trial function for the N -fermion Hamiltonian, H ,
with one- and two-body terms, then $e_0 \leq \langle \psi, H\psi \rangle = E(K)$, where e_0 is the ground-state
energy, K is the one-body reduced density matrix of ψ , and $E(K)$ is the well-known
expression in terms of direct and exchange energies. If an *arbitrary* one-body K is
given, which does not come from a determinantal ψ , then $E(K) \geq e_0$ does not necessarily
hold. It is shown, however, that if the two-body part of H is positive, then in fact e_0
 $\leq e_{\text{HF}} \leq E(K)$, where e_{HF} is the Hartree-Fock ground-state energy.

Considerable distortions of the correct forms
in the original manuscript have necessitated the
following changes in the printed version:

On page 457, first column, line 5 should read
"... which *satisfies the Pauli principle* is re-
quired."

On page 458, second column, lines 5 and 6
should read "... N orthonormal vectors $V^1, \dots$,
 V^N in l^2 , the Hilbert space of square-summable
sequences indexed by the positive integers (i.e.,
 $V \in l^2$ means $\sum_{j=1}^{\infty} |V_j|^2 < \infty$), such that $\sum_{i=1}^N |V_j^i|^2$
 $= c_j$."

On page 458, second column, lines 5 and 6 in
the proof of the theorem should read "Let $V^1, \dots$,
 V^N be the vectors of the lemma ..."

On page 458, second column, line 11 in the
proof of the theorem should read "... over
 $[0, 2\pi)^{2+} \dots$ "

In addition, the following correction should be
noted:

On page 457, the c_j on the right-hand side of
Eq. (3) should be deleted.

Part IV

Thomas-Fermi and Related Theories

Thomas-fermi and related theories of atoms and molecules*

Elliott H. Lieb

Departments of Mathematics and Physics, Princeton University, POB 708, Princeton, New Jersey 08544

This article is a summary of what is known rigorously about Thomas-Fermi (TF) theory with and without the Dirac and von Weizsäcker corrections. It is also shown that TF theory agrees asymptotically, in a certain sense, with nonrelativistic quantum theory as the nuclear charge z tends to infinity. The von Weizsäcker correction is shown to correct certain undesirable features of TF theory and to yield a theory in much better agreement with what is believed (but as yet unproved) to be the structure of real atoms. Many open problems in the theory are presented.

CONTENTS

I. Introduction	603
II. Thomas-Fermi Theory	604
A. The definitions of Thomas-Fermi and related theories	604
B. Domain of definition of the energy functional	604
C. Minimization of the energy functional	605
D. The Thomas-Fermi equation and properties of the density	606
E. The virial and related theorems	608
F. The Thomas-Fermi theory of solids	608
G. The Thomas-Fermi theory of screening	609
H. The Firsov variational principle	610
III. The "No-Binding" and Related Potential-Theoretic Theorems	611
A. Some variational principles and Teller's lemma	611
B. The case of flat j' (TFD)	614
C. No-binding theorems	614
IV. Dependence of the Thomas-Fermi Energy on the Nuclear Coordinates	615
A. The many-body potentials	615
B. The positivity of the pressure	616
C. The long-range interaction of atoms	619
V. Thomas-Fermi Theory as the $Z \rightarrow \infty$ Limit of Quantum Theory	620
A. The $Z \rightarrow \infty$ limit for the energy and density	620
1. Upper bound for E_N^Q	621
2. Lower bound for E_N^Q	621
3. Correlation functions	622
B. The Scott conjecture for the leading correction	623
C. A picture of a heavy atom	624
VI. Thomas-Fermi-Dirac Theory	625
A. The TFD minimization problem	625
B. The j model	625
C. The relation of the j model to TFD theory	627
VII. Thomas-Fermi-von Weizsäcker Theory	628
A. The TFW minimization problem	628
B. Properties of the density and λ_c	631
C. Binding in TFW theory	634
D. The Z^2 correction and the behavior near the nuclei	635
VIII. Thomas-Fermi-Dirac-von Weizsäcker Theory	638
Acknowledgments	639
References	639
Index	641

I. INTRODUCTION

In recent years some of the properties of the Thomas-Fermi (TF) and related theories for the ground states of nonrelativistic atoms and molecules with fixed nuclei have been established in a mathematically rigorous way. The aim of these notes is to summarize that work to date—at least as far as the author's knowledge of the subject goes. In addition, some open problems in the subject will be stated.

TF theory was invented independently by Thomas (1927) and Fermi (1927). The exchange correction was introduced by Dirac (1930), and the gradient correction to the kinetic energy by von Weizsäcker (1935).

No attempt will be made to summarize the voluminous subject of TF theory. Such a summary would have to include many varied applications, many formulations of related theories (e.g., relativistic corrections to TF theory, nonzero temperature TF theory) and reams of data and computations. Some reviews exist (March, 1957; Gombás, 1949; Torrens, 1972), but they are either not complete or not up to date.

We shall concentrate on nonrelativistic TF and related theories for the ground state with the following goals in mind:

(1) The definition of TF and related theories (i.e., the von Weizsäcker and Dirac corrections). The main question here is whether the theories are well defined mathematically and whether the equations to which they give rise have (unique) solutions.

(2) Properties of TF and related theories. It turns out that, unlike the correct Schrödinger, quantum (Q) theory, the TF and related theories have many interesting physical properties that can be deduced without computation. Some of these properties are physically realistic and some are not, e.g., Teller's no-binding theorem. As will be seen, however, the no-binding result is natural and correct if TF theory is placed in its correct physical context as a large- Z (= nuclear-charge) theory.

(3) The relation of TF theory to Q theory. The main result will be that TF theory is exact in the large- Z (nuclear-charge) limit. For this reason, TF theory should be taken seriously as one of the cornerstones of atomic physics. The only other regime in which it is possible to make simple, exact statements is the one-electron hydrogenic atom. The natural open question is to find the leading correction, in Z , beyond TF theory. This will lead to a discussion of the Scott correction

*This article appears in the Proceedings of the NATO Advanced Study Institute on Rigorous Atomic and Molecular Physics held at Erice in June, 1980, edited by G. Velo and A. S. Wightman, and published by Plenum Corporation. The present Rev. Mod. Phys. version contains corrections of some errors in the Plenum version.

(Scott, 1952) which, while it is very plausible, has not yet been proved. It turns out that Thomas-Fermi-von Weizsäcker (TFW) theory has precisely the properties that Scott predicts for Q theory. Moreover, TFW theory remedies some defects of TF theory: It displays atomic binding, it gives exponential falloff of the density at large distances, it yields a finite density at the nucleus, and negative ions are stable (i.e., bound).

The work reported here originated in articles by Lieb and Simon, 1973 and 1977 (hereafter LS). Subsequently, the ideas were developed by, and in collaboration with, Benguria and Brezis. I am deeply indebted to these co-workers.

Since many unsolved problems remain, these notes are more in the nature of a progress report than a textbook. The proofs of many theorems are sketchy, or even absent, but it is hoped that the interested reader can fill in the details with the help of the references. Unless clearly stated otherwise, however, everything presented here is meant to be rigorous.

II. THOMAS-FERMI THEORY

The theories will be stated in this section purely as mathematical problems. Their physical motivation from Q theory will be explained in Sec. V. In order to present the basic ideas as clearly as possible, only TF theory will be treated in this section; the variants will be treated in Secs. VI, VII and VIII. However, the basic definitions of all the theories will be given in Sec. II.A, and there will be some mention of Thomas-Fermi-Dirac (TFD) theory in Sec. II.B and Sec. III.

A. The definitions of Thomas-Fermi and related theories

All the theories we shall be concerned with start with some *energy functional* $\mathcal{E}(\rho)$, where ρ is a non-negative function on three-space, $\mathbb{R}^3$. ρ is called a *density* and physically is supposed to be the electron density in an atom or molecule.

The functionals will involve the following function V and constant U :

$$V(x) = \sum_{j=1}^k z_j |x - R_j|^{-1}, \quad (2.1)$$

$$U = \sum_{1 \leq i < j \leq k} z_i z_j |R_i - R_j|^{-1}. \quad (2.2)$$

$V(x)$ is the electrostatic potential of k nuclei of charges (in units in which the electron charge $e = -1$) $z_1, \dots, z_k > 0$, and located at $R_1, \dots, R_k \in \mathbb{R}^3$. The R_i are distinct. The positivity of the z_i is important for many of the theorems; while TF theory makes mathematical sense when some $z_i < 0$, it has not been investigated very much in that case. U is the repulsive electrostatic energy of the nuclei.

TF-type theories can, of course, be defined for potentials that are not Coulombic, but many of the interesting properties presented here rely on potential theory and hence will not hold for non-Coulombic potentials. This is discussed in Sec. III. There is, however, one generalization of Eqs. (2.1) and (2.2) that can be made without spoiling the theory, namely, that the nuclei can be "smeared out," i.e., the following replacements can be

made:

$$z_j |x - R_j|^{-1} \rightarrow \int dm_j(y) |x - R_j + y|^{-1}, \quad (2.3)$$

$$z_i z_j |R_i - R_j|^{-1} \rightarrow \int dm_i(y) dm_j(w) |y - w - R_i + R_j|^{-1}, \quad (2.4)$$

where m_i is a positive measure (not necessarily spherically symmetric) of mass z_i .

The functional for TF theory is

$$\mathcal{E}(\rho) = \frac{3}{5} \gamma \int \rho(x)^{5/3} dx - \int \rho(x) V(x) dx + D(\rho, \rho) + U, \quad (2.5)$$

where

$$D(g, f) = \frac{1}{2} \iint g(x) f(y) |x - y|^{-1} dx dy.$$

All integrals are three dimensional.

γ is an arbitrary positive constant, but to establish contact with Q theory we must choose

$$\gamma_p = (6\pi^2)^{2/3} \hbar^2 (2mq^{2/3})^{-1}, \quad (2.6)$$

where $\hbar = h/2\pi$, h = Planck's constant, and m is the electron mass. q is the number of spin states ($= 2$ for electrons).

U appears in $\mathcal{E}$ as a constant, ρ -independent term. It is unimportant for the problem of minimizing $\mathcal{E}$ with respect to ρ . Nevertheless U will be very important when we consider how the minimum depends on the R_i , e.g., in the no-binding theorem (Sec. III.C).

For the Thomas-Fermi-Dirac (TFD) theory

$$\mathcal{E}^{\text{TFD}}(\rho) = \mathcal{E}(\rho) - \frac{3}{4} C_e \int \rho(x)^{4/3} dx, \quad (2.7)$$

with C_e a positive constant. In the original theory (Dirac, 1930), the value $C_e = (6/\pi q)^{1/3}$ was used for reasons which will be explained in Sec. VI. This value is not sacrosanct, however, and it is best to leave C_e as an adjustable constant.

The Thomas-Fermi-von Weizsäcker theory (TFW) is given by (von Weizsäcker, 1935)

$$\mathcal{E}^{\text{TFW}}(\rho) = \mathcal{E}(\rho) + \delta \int [(\nabla \rho^{1/2})(x)]^2 dx, \quad (2.8)$$

with $\delta = A\hbar^2/2m$, and A an adjustable constant. Originally, A was taken to be unity, but in Sec. VII.D it will be seen that $A = 0.186$ is optimum from one point of view.

The most complicated, and least analyzed, case is the combination of all three (Sec. VIII):

$$\begin{aligned} \mathcal{E}^{\text{TFDW}}(\rho) = & \mathcal{E}(\rho) - \frac{3}{4} C_e \int \rho(x)^{4/3} dx \\ & + \delta \int [(\nabla \rho^{1/2})(x)]^2 dx. \end{aligned} \quad (2.9)$$

The first question to face is the following.

B. Domain of definition of the energy functional

Since ρ is supposed to be the electron density we require $\rho(x) \geq 0$ and

$$\int \rho(x) dx = \lambda = \text{electron number} \quad (2.10)$$

is finite. In addition we require $\rho \in L^{5/3}$ in order that the first term in $\mathcal{E}(\rho)$ (called the kinetic energy term) be finite. λ is not necessarily an integer.

Definition. A function f is said to be in L^p if $[\int |f(x)|^p dx]^{1/p} = \|f\|_p$ is finite, $1 \leq p < \infty$. $\|f\|_\infty = \text{ess sup } |f(x)|$ (see Theorem 3.12).

If $f \in L^p \cap L^q$ with $p < q$ then $f \in L^t$ for all $p < t < q$.
 $\|f\|_t \leq \|f\|_p^{p/t} \|f\|_q^{q/t}$, where $\lambda p^{-1} + (1-\lambda)q^{-1} = t^{-1}$.

Proposition 2.1. If $\rho \in L^{5/3} \cap L^1$ then all the terms in $\mathcal{E}$ and $\mathcal{E}^{\text{TFD}}$ are finite. If $\int \rho \leq \lambda$ then $\mathcal{E}(\rho)$ and $\mathcal{E}^{\text{TFD}}(\rho)$ are bounded below by some constant $C(\lambda)$. Furthermore, for all λ , $\mathcal{E}(\rho) > C > -\infty$ for some fixed C .

Proof. The first part is an easy application of Young's and Hölder's inequalities. The second part requires a slightly more refined estimate of the Coulomb energies (cf. LS). ■

Remark. Although $\mathcal{E}^{\text{TFW}}$ will be seen to be also bounded below by a constant independent of λ , neither $\mathcal{E}^{\text{TFD}}$ nor $\mathcal{E}^{\text{TFDW}}$ is so bounded. This fact leads to an amusing unphysical consequence of the D theories which will be mentioned later.

A very important fact (which, incidentally, is not true for Hartree-Fock theory) is the following.

Proposition 2.2. $\rho \rightarrow \mathcal{E}(\rho)$ is strictly convex, i.e., $\mathcal{E}(\lambda\rho_1 + (1-\lambda)\rho_2) < \lambda\mathcal{E}(\rho_1) + (1-\lambda)\mathcal{E}(\rho_2)$ for $0 < \lambda < 1$ and $\rho_1 \neq \rho_2$.

Proof. ρ^p is strictly convex for $p > 1$. $\int V\rho$ is linear in ρ and hence convex. $D(\rho, \rho)$ is strictly convex since the Coulomb kernel $|x-y|^{-1}$ is positive definite. ■

Remark. $\mathcal{E}^{\text{TFW}}$ is also strictly convex, but the functionals $\mathcal{E}^{\text{TFD}}$ and $\mathcal{E}^{\text{TFDW}}$ are not convex because of the $-\int \rho^{5/3}$ term. However, $\mathcal{E}^{\text{TFD}}$ can be "convexified" in a manner to be described in Sec. VI.

C. Minimization of the energy functional

The central problem is to compute

$$E(\lambda) = \inf \left\{ \mathcal{E}(\rho) \mid \rho \in L^{5/3} \cap L^1, \int \rho = \lambda \right\} \quad (2.11)$$

and

$$e(\lambda) = E(\lambda) - U. \quad (2.12)$$

$E(\lambda)$ is the TF energy for a given electron number, λ , and $e(\lambda)$ is the *electronic contribution to the energy*. The "inf" in Eq. (2.11) is important because, as we shall see, the minimum is not always achieved, although the inf always exists by Prop. 2.1.

Theorem 2.3. $e(\lambda)$ is convex, negative if $\lambda > 0$, nonincreasing and bounded below. Furthermore,

$$E(\lambda) = \inf \left\{ \mathcal{E}(\rho) \mid \rho \in L^{5/3} \cap L^1, \int \rho \leq \lambda \right\}. \quad (2.13)$$

Proof. The first part follows from Prop. 2.2 together with the observation that $V(x) \rightarrow 0$ as $|x| \rightarrow \infty$. This means that if λ increases we can add some $\delta\rho$ arbitrarily far from the origin so that $\mathcal{E}(\rho + \delta\rho) - \mathcal{E}(\rho) < \varepsilon$ for

any $\varepsilon > 0$. Equation (2.13) is a simple consequence of the monotonicity of $e(\lambda)$ and $E(\lambda)$. (cf. LS). ■

Equation (2.13) has an important advantage over (2.11), as Theorem 2.4 shows.

Theorem 2.4. There exists a unique ρ that minimizes $\mathcal{E}(\rho)$ on the set $\int \rho \leq \lambda$.

Note. Uniqueness means, of course, that ρ is determined only almost everywhere (a.e.).

Proof. (See LS.) Since $\mathcal{E}(\rho)$ is strictly convex, a minimum, if there is one, must be unique. Let $\rho^{(n)}$ be a minimizing sequence for $\mathcal{E}$, namely $\mathcal{E}(\rho^{(n)}) \rightarrow E(\lambda)$ and $\int \rho^{(n)} \leq \lambda$. It is easy to see that $\int (\rho^{(n)})^{5/3} \leq c$, where c is some constant; this in fact comes out of the simple estimates used in the proof of Prop. 2.1. We should like to extract a convergent subsequence from the given $\rho^{(n)}$. This cannot be done *a priori* in the strong topology, but the Banach-Alaoglu theorem tells us that a $L^{5/3}$ weakly convergent subsequence can be found; this will be denoted by $\rho^{(n)}$. We should like to prove

$$\liminf \mathcal{E}(\rho^{(n)}) \geq \mathcal{E}(\rho). \quad (2.14)$$

Since $\rho^{(n)} \rightarrow \rho$ weakly in $L^{5/3}$ we have (by the Hahn-Banach theorem, for example) that

$$\liminf \int [\rho^{(n)}]^{5/3} \geq \int \rho^{5/3}, \quad (2.15)$$

$$\liminf D(\rho^{(n)}, \rho^{(n)}) \geq D(\rho, \rho). \quad (2.16)$$

The term $-\int V\rho$ requires slightly more delicate treatment. Write $|x-R|^{-1} = f(x) + g(x)$, where $f(x) = |x-R|^{-1}$ for $|x-R| \leq 1$ and $f(x) = 0$ otherwise. $f \in L^{5/2}$ and $\int \rho^{(n)} f \rightarrow \int \rho f$ by weak convergence. On the other hand, $g \in L^{5/3+\varepsilon}$ for all $\varepsilon > 0$. $\rho^{(n)}$ is bounded in $L^{5/3}$ and in L^1 (by λ), so it is bounded in all L^q with $1 \leq q \leq \frac{5}{3}$ and therefore $\rho^{(n)} \rightarrow \rho$ weakly in L^q as well as in $L^{5/3}$. Fix $\infty > \varepsilon > 0$ and let q be dual to $3+\varepsilon$. Then $\int g\rho^{(n)} \rightarrow \int g\rho$. This proves Eq. (2.14) which, since $E(\lambda) = \liminf \mathcal{E}(\rho^{(n)})$ and $E(\lambda) \leq \mathcal{E}(\rho)$, implies that ρ is minimizing provided we can show $\int \rho \leq \lambda$. This follows from the fact that if $\int \rho > \lambda$ then there is a bounded set A such that $\int_A \rho > \lambda$. If α is the characteristic function of A then $\alpha \in L^{5/2}$ and $\lambda \geq \int \alpha \rho^{(n)} \rightarrow \int \alpha \rho$ by weak $L^{5/3}$ convergence. ■

Remark. The proof of Theorem 2.4 can be considerably shortened by using Mazur's (1933) theorem. $\rho \rightarrow \mathcal{E}(\rho)$ is obviously norm continuous and hence norm lower semicontinuous. Mazur's theorem says that the convexity of $\mathcal{E}(\rho)$ then automatically implies weak lower semicontinuity since norm closed convex sets are automatically weakly closed. The proof given above has the virtue of an explicit demonstration of the weak lower semicontinuity.

Remark. The analogous proof in TFD theory will be harder, since $\rho^{5/3} - \rho^{4/3}$ is not convex, monotone, or positive; hence we cannot say that

$$\liminf \int [\rho^{(n)}]^{5/3} - [\rho^{(n)}]^{4/3} \geq \int \rho^{5/3} - \rho^{4/3}.$$

However, in TFW theory a different strategy, using Fatou's lemma, will be employed to deal with these terms. The strategy also works for TFDW theory. Thus the introduction of the W term (2.8) makes part of the proof easier. It would be desirable to know how to

use Fatou's lemma (which does not require convexity) in the TF and TFD proofs.

Since $E(\lambda)$ is nonincreasing, bounded, and convex (and hence continuous) we can make the following definition in TF theory.

Definition. λ_c , the critical λ , is the largest λ with the property that for all $\lambda' < \lambda$, $E(\lambda') > E(\lambda)$. Equivalently, if $E(\infty) = \lim_{\lambda \rightarrow \infty} E(\lambda)$ then $\lambda_c = \inf \{ \lambda \mid E(\lambda) = E(\infty) \}$. In principle λ_c could be $+\infty$, but this will not be the case. In TFD and TFDW theories $E(\lambda)$ is not bounded and the above definition has to be generalized. λ_c is the largest λ with the property that $2E(\lambda) < E(\lambda - \varepsilon) + E(\lambda + \varepsilon)$ for all $0 < \varepsilon < \lambda$. In other words, $E(\lambda) = E(\lambda_c) + (\text{const})(\lambda - \lambda_c)$ for $\lambda \geq \lambda_c$. The j model in TFD theory is bounded, so the first definition is applicable to that model. λ_c will be shown to be $Z = \sum z_j$ in TF and TFD theory, Theorem 3.18. In TFW theory, $\lambda_c > Z$ (Theorem 7.19).

Theorems 2.3, 2.4, and Proposition 2.2 yield the following picture of the minimization problem in TF theory.

Theorem 2.5. For $\lambda \leq \lambda_c$ there exists a unique minimizing ρ with $\int \rho = \lambda$. On the set $[0, \lambda_c]$, $E(\lambda)$ is strictly convex and monotone decreasing. For $\lambda > \lambda_c$ there is no minimizing ρ with $\int \rho = \lambda$, and $E(\lambda) = E(\lambda_c)$; the minimizing ρ in Theorem 2.4 is the ρ for λ_c .

Proof. For $\lambda \leq \lambda_c$ use the ρ given by Theorem 2.4 and note that if $\lambda' = \int \rho < \lambda$ then $E(\lambda') = \mathcal{G}(\rho) = E(\lambda)$. The strict convexity is trivial: if $\lambda = a\lambda_1 + (1-a)\lambda_2$ use $a\rho_1 + (1-a)\rho_2$ as a trial function for λ . On the other hand, for $\lambda > \lambda_c$ the ρ given by Theorem 2.4 will have $\int \rho = \lambda_c$ because if a minimum existed with $\int \rho = \lambda' > \lambda_c$ then $\tilde{\rho} \equiv (\rho + \rho_c)/2$ (with ρ_c being the ρ for λ_c) would satisfy $\lambda_c < \int \tilde{\rho} = \frac{1}{2}(\lambda' + \lambda_c) < \lambda'$ but, by strict convexity,

$$\mathcal{G}(\tilde{\rho}) < [\mathcal{G}(\rho) + \mathcal{G}(\rho_c)]/2 = E(\lambda_c),$$

which is a contradiction. ■

The general situation is shown in Fig. 1. There Z is shown as less than λ_c , while that is the case for TFW theory, in TF and TFD theory $\lambda_c = Z$. The straight portion to the right of λ_c is horizontal for TF and TFW, but has a negative slope for TFD and TFDW. The slope at the origin is infinite for TF and TFD but finite for TFW and TFDW.

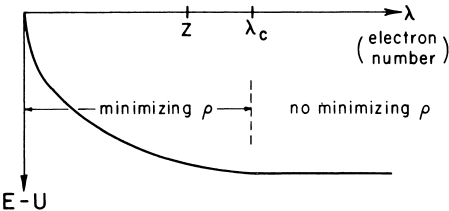


FIG. 1. "The electronic part" of the TF energy, $E-U$, is shown schematically as a function of the "electron number", $\lambda = \int \rho$. For $\lambda \leq \lambda_c$ there is a unique ρ that minimizes the TF energy $\mathcal{G}(\rho)$. For $\lambda > \lambda_c$ there is no such ρ . In TF theory $\lambda_c = Z = \sum_{j=1}^J z_j$ = total nuclear charge. $E-U$ is constant for $\lambda \geq \lambda_c$. These features are different for TF, TFD, and TFDW theories (see text).

D. The Thomas-Fermi equation and properties of the density

The variational derivative of $\mathcal{G}(\rho)$ is $\delta \mathcal{G} / \delta \rho = \gamma \rho^{2/3}(x) - \phi_\rho(x)$ where

$$\phi_\rho(x) = V(x) - \int \rho(y) |x-y|^{-1} dy. \quad (2.17)$$

A Lagrange multiplier μ should be added to $\delta \mathcal{G} / \delta \rho$ to insure that $\int \rho = \lambda$. It is then expected that $\delta \mathcal{G} / \delta \rho + \mu = 0$ if $\rho(x) > 0$, but $\delta \mathcal{G} / \delta \rho + \mu \geq 0$ if $\rho(x) = 0$ because negative variations of $\rho(x)$ are not allowed. The two situations can be written as

$$\gamma \rho^{2/3}(x) = \max[\phi_\rho(x) - \mu, 0] \equiv [\phi_\rho(x) - \mu]_+. \quad (2.18)$$

This is the **TF equation**. (Note that the $[\]_+$ is very important.) This formal manipulation is, indeed, correct.

Theorem 2.6. If ρ minimizes $\mathcal{G}(\rho)$ with $\int \rho = \lambda \leq \lambda_c$ then ρ satisfies Eq. (2.18) for some (unique) $\mu(\lambda)$. Conversely if ρ, μ satisfy Eq. (2.18) and $\rho \in L^1 \cap L^{5/3}$ then ρ minimizes $\mathcal{G}(\rho)$ for $\lambda = \int \rho$. Hence (2.18) can have at most one solution ρ, μ with $\int \rho = \lambda$. If $\lambda = \lambda_c$ then $\mu = 0$.

Proof. The first part is standard in the calculus of variations. Now let $\rho_i, \mu_i, i=1,2$, satisfy Eq. (2.18) with the same λ . Let $F_i(h) = (3\gamma/5) \int h^{5/3} - \int \phi_i h$. It is easy to check that $F_i(h)$ has a unique minimum, F_i , on the set $\int h = \lambda, h \geq 0$; the minimizing h_i is ρ_i . However, $F_1(\rho_2) + F_2(\rho_1) = F_1 + F_2 - D(\rho_1 - \rho_2, \rho_1 - \rho_2)$. This is a contradiction unless $\rho_1 = \rho_2$ (and hence $\mu_1 = \mu_2$). The last part (i.e., $\mu = 0$) follows by considering the absolute minimum of $\mathcal{G}(\rho)$, in which case no μ is necessary. But this is equivalent to setting $\mu = 0$. This minimum occurs for $\lambda \geq \lambda_c$ but as we have shown, only at λ_c is there a minimizing ρ (cf. LS). ■

Remarks. In Sec. III a proof of the uniqueness part of Theorem 2.6 which uses only potential theory will be given. It should be noted that we arrived at the existence of a solution to Eq. (2.18) by first considering the minimization problem. A direct attack on (2.18) is rather difficult. Such a direct approach was carried out by Hille (1969) in the atomic case, but even in that case he did not prove that the spherically symmetric solution is the only one; our uniqueness result guarantees that.

Theorem 2.7. $E(\lambda)$ is continuously differentiable and $dE/d\lambda = -\mu(\lambda)$ if $\lambda < \lambda_c$. $dE/d\lambda = 0$ if $\lambda \geq \lambda_c$. Thus $-\mu(\lambda)$ is the chemical potential.

Proof. The convexity and boundedness of $\mathcal{G}(\rho)$ is used. (See LS, Theorem II.10 and Lemma II.27.) ■

It will be noted that we have not used the fact that V is Coulombic, only that it vanishes at ∞ . Likewise, the only property of the kernel $|x-y|^{-1}$ that was used was its positive definiteness. In Sec. III we shall exploit the fact that $|x-y|^{-1}$ is Coulombic and, to a lesser extent, the fact that V is superharmonic. Also, it will be shown that $\lambda_c = Z = \sum_{j=1}^J z_j$.

Definition. A function $f(x)$ defined on an open set $\Omega \subset \mathbb{R}^3$ is **superharmonic** on Ω if, for almost all $x \in \Omega$ and for almost all spheres centered at x , but contained in Ω , $f(x) \geq (\text{the average of } f \text{ on the sphere})$, i.e., $f(x) \geq (4\pi)^{-1} \times \int_{|y-x|=R} f(x+y) dy$. This is the same as $\Delta f \leq 0$ (in the

sense of distributions) in Ω . f is *subharmonic* if $-f$ is superharmonic. f is *harmonic* if it is both subharmonic and superharmonic.

In Sec. III potential theory will shed considerable light on the solution to Eq. (2.18). Here we shall concentrate on some other aspects of (2.18).

Let us assume that $V(x) = \sum z_j |x - R_j|^{-1}$. ϕ denotes ϕ_ρ for the solution to Eq. (2.18). In Sec. III we show $\phi(x) > 0$. As a distribution,

$$\begin{aligned} -\Delta \phi(x)/4\pi &= \sum z_j \delta(x - R_j) - \rho(x) \\ &= \sum z_j \delta(x - R_j) - \gamma^{-3/2} (\phi(x) - \mu)^{3/2}. \end{aligned} \quad (2.19)$$

This is the TF *differential equation* and is *equivalent* to Eq. (2.18). It involves ϕ alone. Since $\rho \in L^{5/3} \cap L^1$, ϕ is continuous away from the R_j (Lemma 3.1) and goes to zero as $|x| \rightarrow \infty$. The fact that ϕ goes to zero at infinity is understood as a boundary condition in Eq. (2.19).

Theorem 2.8 (LS Theorem IV.5).

- (a) Near each R_j
 $\rho(x) = (z_j/\gamma)^{3/2} |x - R_j|^{-3/2} + O(|x - R_j|^{-1/2})$
 (b) $\rho(x) \rightarrow 0$ as $|x| \rightarrow \infty$.
 (c) ρ and ϕ are real analytic on $A = \{x | x \neq R_j \text{ all } j, \rho(x) > 0\}$.
 (d) In the neutral case ($\mu = 0$) $\rho(x) > 0$, all x .
 (e) In the ionic case ($\lambda < Z, \mu > 0$) ρ has compact support and ρ and ϕ are C^1 away from the R_j .

Proof. (a) and (b) follow directly from Eq. (2.18). ϕ continuous $\Rightarrow \rho$ continuous away from the $R_j \Rightarrow \phi$ is C^1 everywhere. Then ρ is C^1 away from the R_j . [Note $(\phi - \mu)^{3/2}$ is C^1 if ϕ is C^1]. By a bootstrap argument ρ is C^∞ on A . By Theorem 5.8.6 in Morrey (1966), ϕ is real analytic away from the R_j and where $\phi > \mu$, namely A . Finally, since $\phi(x) \rightarrow 0$ as $|x| \rightarrow \infty$, ρ has compact support when $\mu > 0$. The positivity of ϕ is established in Sec. III, so $\rho > 0$ in the neutral case. ■

In the ionic case ($\lambda < Z$) the set $\Omega = \{x | \rho(x) > 0\}$ is bounded. What can be said about its boundary, $\partial\Omega$? In the atomic case $\partial\Omega$ is, of course, a sphere. In the general case, the TF equation (2.19) is a "free boundary problem" about which Caffarelli and Friedman (1979) have proved the following result among others.

Theorem 2.9. Consider the generalized TF problem with $\frac{5}{3}$ replaced by p and $\frac{3}{2} < p < 2$. There are at most a finite number of open C^1 curves $\Gamma_1, \dots, \Gamma_l$ such that $\partial\Omega \setminus \bigcup_{i=1}^l \Gamma_i$ is a $C^{3+\alpha}$ manifold with $\alpha = (2-p)/(p-1)$.

The next question to consider is the asymptotics of ρ , in the neutral case ($\mu = 0$), as $|x| \rightarrow \infty$. This involves finding universal bounds on ρ . The function $\psi(x) = \gamma^2(3/\pi)^2 |x|^{-4}$ satisfies Eq. (2.19) for $|x| > 0$ and $x \neq R_j$. It is the only "power law" that does so. This was noted by Sommerfeld, who concluded that $\psi(x)$ is the asymptotic form of ϕ . Hille (1969), who was possibly the first to make a serious mathematical study of the TF equation, proved this asymptotic law in the atomic case. It is remarkable that ψ , the asymptotic form of ϕ , is independent of z , and it is just as remarkable that the same form holds

even in the molecular case.

Theorem 2.10 (LS Sec. V.2). Suppose $\mu = 0$ and $|R_j| < R$, for all j and some R . For $r = |x| > R$ let $\phi_+(r)$ (resp. $\phi_-(r)$) be the max (resp. min) of $\phi(x)$ on $|x| = r$ and $C_\pm(r) = \phi_\pm(r)/\psi(r)$ with $\psi(r) = \gamma^2(3/\pi)^2 r^{-4}$. Then $C_\pm(r) \rightarrow 1$ as $r \rightarrow \infty$. Furthermore, if $R < r$ then

- (i) $C_+(R) \geq 1 \Rightarrow C_+(r) \leq C_+(R)$,
 (ii) $C_-(R) \leq 1 \Rightarrow C_-(r) \leq 1$,
 (iii) $C_-(R) \leq 1 \Rightarrow C_-(r) \geq C_-(R)$,
 (iv) $C_-(R) \geq 1 \Rightarrow C_-(r) \geq 1$.

Proof. If f, g are continuous, positive functions on $|x| \geq R$ which go to zero as $|x| \rightarrow \infty$, and if $\gamma^{3/2} \Delta f \leq 4\pi f^{3/2}$, $\gamma^{3/2} \Delta g \geq 4\pi g^{3/2}$ for $|x| > R$, and if $f(x) \geq g(x)$ for $|x| = R$, then $f(x) \geq g(x)$ for all $|x| > R$. This is easily proved by a "maximum argument" as in Sec. III. ϕ is of this type with $\gamma^{3/2} \Delta \phi = 4\pi \phi^{3/2}$. If $C_+(R) \geq 1$, compare $\phi(x)$ with $C_+(R)\psi(x)$. Then $C_+(r) \leq C_+(R)$, all $r > R$. This proves (i) and similarly (iii). To prove (ii) and (iv) compare ϕ with ψ . It remains to show that $C_\pm(r) \rightarrow 1$. C_+ is continuous. We shall show that $\limsup C_+(r) \leq 1$; by a similar argument $\liminf C_-(r) \geq 1$. This will complete the proof. If $C_+(R) > 1$, let $R_0 = \sup\{r | C_+(r) \geq 1\}$ whence $C_+(r) \geq 1$ for $R \leq r \leq R_0$ and $C_+(r) \leq 1$ for $r \geq R_0$. It is then only necessary to consider $R_0 = \infty$. Then, since $C_+(r)$ is decreasing, $C(\infty) = \lim C_+(r)$ exists. Assume $C(\infty) > 1$. Pick $\varepsilon > 0$ and choose R_1 so that $C_+(R_1) < C(\infty) + \varepsilon$. Consider

$$f(x) = \gamma^2(3/\pi)^2(1 + 2b/3)^2 \{ |x| - bR_1 \}^{-4}$$

for $|x| \geq R_1$, and $b < 1$. $\gamma^{3/2} \Delta f \leq 4\pi f^{3/2}$. Choose $b < 1$ such that $(1 + 2b/3)^2 = C_+(R_1)(1 - b)^4$. Then $f \geq \phi$ for $|x| > R_1$ since $f \geq \phi$ when $|x| = R_1$. But this means $C(\infty) \leq (1 + 2b/3)^2 = C(\infty) + \varepsilon$ (1 - b)⁴. Since $b > 0$ satisfying $(1 + 2b/3)^2 = C(\infty)(1 - b)^4$, and ε is arbitrary, $C(\infty) \leq 1$. For the C_- problem use $g(x) = \gamma^2(3/\pi)^2(1 - 2b/3)^2 \{ |x| + bR_1 \}^{-4}$. ■

There are some interesting facts about the possible singularities of TF-type differential equations in a ball. These are related to and complement Theorem 2.10.

Theorem 2.11. Let $B = \{x | 0 < |x| < R\}$ and suppose ϕ satisfies $\Delta \phi(x) = G(\phi(x))$ in the sense of distributions on B , and G is C^1 . Then

- (a) If $\phi \in L^\infty_{\text{loc}}(B)$ and G satisfies $G(t) > \varepsilon t^3$ as $t \rightarrow \infty$, $G(t) < \varepsilon t^3$ as $t \rightarrow -\infty$, $\varepsilon > 0$, there exists a C^2 function on $0 \leq |x| < R$ which agrees with ϕ a.e. in B . Any singularity is thus removable. In particular $-\Delta \phi + \phi^3 = \delta(x)$ has no solution in $B' = \{x | 0 \leq x < R\}$.
 (b) If $\phi \in C^2(B)$, $\phi > 0$, and $G(t) = t^q$, $1 < q < 3$, one of the following is true:

- (i) ϕ has a C^2 extension on B' .
 (ii) $\phi(x) \sim C |x|^{-1}$ as $|x| \rightarrow 0$, $C > 0$ arbitrary
 (iii) $\phi(x) \sim t^{-1} |x|^{-a}$ as $|x| \rightarrow 0$ with $a = 2/(q-1)$, $l^{q-1} = a(a-1)$. This is called the "strong singularity."
 (c) Let $q > 1$ and $B' = \{x | 0 \leq |x| < R\}$. Let $\phi \in L^\infty_{\text{loc}}(B')$ satisfy $\Delta \phi = |\phi|^{q-1} \phi$ in B' in the sense of distributions. There is a universal constant $C_q < \infty$ such that $|\phi(0)| \leq C_q R^{-2/(q-1)}$. This implies that if $\phi \in L^\infty_{\text{loc}}(B)$ satisfies this equation in B , then $|\phi(x)| \leq C_q |x|^{-2/(q-1)}$ for $2|x| < R$. A stronger bound than this is given by Veron (1979) and Brezis and Veron (1980) for $1 < q < 3$.

Proof. (a) is given in Brezis and Veron, 1980, and (b) and (c) are given in Veron, 1979. (c) was given earlier

for $q = \frac{3}{2}$ in Brezis and Lieb, 1979. ■

There are other theorems of this type in Veron, 1979 and Brezis and Lieb, 1979.

See Sec. IV.C for an application of the strong singularity.

There is another property of ρ which can be derived directly from the variational principle, namely,

Theorem 2.12. *In the atomic case $\rho(x)$ is a symmetric, decreasing function.*

Proof. Assume the nucleus is at the origin and let ρ^* be the symmetric, decreasing rearrangement of ρ (for a definition see Lieb, 1977). We claim that if $\rho \neq \rho^*$ then $\mathcal{E}(\rho^*) < \mathcal{E}(\rho)$, thereby proving the theorem. $\int (\rho^*)^q = \int \rho^q$, all ρ . For the Coulomb terms note that when $\int \rho \leq z$ then $f(x) = z|x|^{-1} - |x|^{-1} * (\rho^*)$ is a symmetric, decreasing function; hence $\int f\rho \leq \int f\rho^*$. Thus $P(\rho) \equiv D(\rho, \rho) - \int V\rho = D(\rho - \rho^*, \rho - \rho^*) - \int \rho f - D(\rho^*, \rho^*)$ and thus $P(\rho) > P(\rho^*)$ if $\rho \neq \rho^*$. ■

Notation. $f * g$ denotes convolution, namely $(f * g)(x) \equiv \int f(x-y)g(y)dy$.

Remarks. (i) The same theorem (and proof) holds for the TFD, TFW, and TFDW theories provided $\lambda = \int \rho \leq z$. The only additional fact needed for the W theories is that $\int (\nabla \psi)^2 \geq \int (\nabla \psi^*)^2$ (see Lieb, 1977, appendix). In fact, Theorem 2.12 holds for all λ in TFW theory (Theorem 7.26).

(ii) The spherically symmetric (but not the decreasing) property of ρ also follows from the uniqueness of ρ which, in turn, follows from the strict convexity of $\mathcal{E}$. The decreasing property also follows from Eq. (2.18) since ϕ is decreasing by Newton's theorem.

E. The virial and related theorems

Let us generalize the TF functional $\mathcal{E}$ by multiplying the term $D(\rho, \rho)$ in Eq. (2.5) by a parameter $\beta > 0$. $e(\lambda) = E(\lambda) - U$ in Eq. (2.12) is then a function of γ , $\{z_j\}$, and β . Define

$$K = \frac{3}{5} \gamma \int \rho^{5/3}, \quad R = \beta D(\rho, \rho), \quad A = \int V\rho, \quad (2.20)$$

with ρ being the minimizing ρ for $\int \rho = \lambda$ with $\lambda \leq \lambda_c$. [By scaling, $\lambda_c(\beta) = \lambda_c(\beta=1)/\beta$.]

Theorem 2.13. *$e(\lambda, \gamma, \{z_j\}, \beta)$ is a C^1 function of its $k+3$ arguments (assuming all are >0 , except for β which is ≥ 0 , and $\lambda \leq \lambda_c$). e is convex in λ and jointly concave in $(\gamma, \{z_j\}, \beta)$. Moreover, $\partial e / \partial \gamma = K/\gamma$, $\partial e / \partial \beta = R/\beta$, $\partial e / \partial \lambda = -\mu$, $\partial e / \partial z_j = -\int \rho(x) |x - R_j|^{-1} dx$. This implies*

$$\partial E / \partial z_j = \lim_{x \rightarrow R_j} \{ \phi(x) - z_j |x - R_j|^{-1} \}. \quad (2.21)$$

Proof. See LS. The proof uses the convexity of $\rho \rightarrow \mathcal{E}(\rho)$. The concavity in the parameters is a trivial consequence of the variational principle and the linearity of $\mathcal{E}$ in the parameters. ■

Now we return to $\beta=1$.

Theorem 2.14. (a) $5K/3 = A - 2R - \mu\lambda$,
(b) *for an atom ($k=1$), $2K=A-R$.*

Proof. (a) Simply multiply the TF equation (2.18) by ρ

and integrate. Alternatively, note that ρ minimizes $G(\rho) = \mathcal{E}(\rho) + \mu \int \rho$ on all of $L^{5/3} \cap L^1$. Therefore $f(t) \equiv G(\rho_t)$, with $\rho_t(x) = t\rho(x)$, has its minimum at $t=1$. But $df/dt=0$ gives (a).

(b) Here, scaling is essential. Consider $\rho_t(x) = t^3 \rho(tx)$, so that $\int \rho_t = \lambda$. Then $f(t) = \mathcal{E}(\rho_t)$ has its minimum at $t=1$ and $df/dt=0$ gives (b). ■

Remark. (b) is called the *Virial theorem*. *A priori* there is an analog of (b) for a molecule. Suppose that, with λ fixed, e is stationary with respect to all R_j , i.e., $\nabla_{R_j} e = 0$. Then, by the same scaling argument together with $R_j \rightarrow tR_j$, one would conclude that $2K = A - R - U$, equivalently $K + E = 0$. See Fock, 1932 and Jensen, 1933. The difficulty with this is that there are no stationary points for $k \geq 2$. The no-binding Theorem 3.23 shows that there are no global minima, and the positivity of the pressure proved in Sec. IV.B shows that there are no local minima (at least for neutral molecules). There it will be shown that for $k \geq 2$, the pressure P satisfies

$$3P = K + E > 0 \text{ for neutral molecules.} \quad (2.22)$$

For non-neutral molecules, a sharpening of Theorem 4.7 into a strict inequality for the derivative would suffice to show the absence of local minima.

For a neutral atom, (a) and (b) combine to give the following simple ratios:

$$R : K : -e : A = 1 : 3 : 3 : 7. \quad (2.23)$$

The energy of a neutral atom is

$$e = E = -3.678 \, 74 z^{7/3} / \gamma.$$

I thank D. Liberman for this numerical value.

Scaling. Suppose the nuclear coordinates R_i are replaced by lR_i with $l > 0$. If $\underline{z}, \underline{R}$ denote the nuclear charges and coordinates, and if $E(\underline{z}, \lambda, l\underline{R})$, $-\mu(\underline{z}, \lambda, l\underline{R})$, $\rho(\underline{z}, \lambda, l\underline{R}; x)$, and $\phi(\underline{z}, \lambda, l\underline{R}; x)$ denote the TF energy, chemical potential, density, and potential with $\int \rho = \lambda$, then

$$\begin{aligned} E(\underline{z}, \lambda, l\underline{R}) &= l^{-7} E(l^3 \underline{z}, l^3 \lambda, \underline{R}), \\ \mu(\underline{z}, \lambda, l\underline{R}) &= l^{-4} \mu(l^3 \underline{z}, l^3 \lambda, \underline{R}), \\ \rho(\underline{z}, \lambda, l\underline{R}; x) &= l^{-6} \rho(l^3 \underline{z}, l^3 \lambda, \underline{R}; l^{-1}x), \\ \phi(\underline{z}, \lambda, l\underline{R}; x) &= l^{-4} \phi(l^3 \underline{z}, l^3 \lambda, \underline{R}; l^{-1}x). \end{aligned} \quad (2.24)$$

This is a trivial consequence of the scaling properties of $\mathcal{E}(\rho)$.

F. The Thomas-Fermi theory of solids

A solid is viewed as a large molecule with the nuclei arranged periodically. For simplicity, but not necessity, let us suppose that there is one nucleus of charge z per unit cell located on the points of $\mathbb{Z}^3 \subset \mathbb{R}^3$. ($\mathbb{Z}^3$ consists of the points with integer coordinates.) If Λ is a finite subset of $\mathbb{Z}^3$ we want to know if, as $\Lambda \rightarrow \infty$ in a suitable sense, the energy/unit volume $|\Lambda|^{-1} E_\Lambda$ has a limit E , and ρ_Λ has a limit ρ , which is a periodic function. Here, $|\Lambda|$ is the volume of Λ . If so, the equation for ρ and an expression for E in terms of ρ is required. Naturally, it is necessary to consider only *neutral* systems, for otherwise $|\Lambda|^{-1} E_\Lambda \rightarrow \infty$. Everything works out as expected except for one mildly surprising thing; a

quantity ψ_0 appears in the equation for ρ which, while it looks like a chemical potential, and is often assumed to be one, is not a chemical potential. ψ_0 is the average electric potential in the solid. All of this is proved in LS, Sec. VI.

Definition. A sequence of domains $\{\Lambda_i\}$ in Z^3 is said to *tend to infinity* (denoted by $\Lambda \rightarrow \infty$) if

(i) $\cup_{i=1}^{\infty} \Lambda_i = Z^3$, (ii) $\Lambda_{i+1} \supset \Lambda_i$, (iii) $\Lambda_i^c \subset Z^3$ is the set of points not in Λ_i , but whose distance to Λ_i is less than h . Then $|\Lambda_i^c|/|\Lambda_i| \rightarrow 0$ for each $h > 0$. $\Gamma = \{x \in R^3 \mid |x| < \frac{1}{2}\}$ is the elementary cube centered at the origin.

Theorem 2.15. As $\Lambda \rightarrow \infty$ the following limits exist and are independent of the sequence Λ_i :

$$(i) \phi(x) = \lim_{\Lambda \rightarrow \infty} \phi_{\Lambda}(x).$$

ϕ is periodic, $\gamma\rho(x)^{2/3} \equiv \phi(x)$, and the convergence is uniform on compacts in R^3 .

$$(ii) \phi(x) = \lim_{\Lambda \rightarrow \infty} |\Lambda|^{-1} \sum_{y \in \Lambda} \phi_{\Lambda}(x+y),$$

$$(iii) \lim_{x \rightarrow 0} \phi(x) - z|x|^{-1} = \lim_{\Lambda \rightarrow \infty} |\Lambda|^{-1} \sum_{y \in \Lambda} \lim_{x \rightarrow y} \phi_{\Lambda}(x) - z|x-y|^{-1},$$

$$(iv) \int_{\Gamma} \rho = \lim_{\Lambda \rightarrow \infty} \int_{\Gamma} \rho_{\Lambda} = z,$$

$$(v) \int_{\Gamma} \rho^{5/3} = \lim_{\Lambda \rightarrow \infty} |\Lambda|^{-1} \int_{\Gamma} \rho_{\Lambda}^{5/3},$$

$$(vi) E = \lim_{\Lambda \rightarrow \infty} |\Lambda|^{-1} E_{\Lambda}.$$

Definition. $G(x)$ is the periodic Coulomb potential. It is defined up to an unimportant additive constant in Γ by $-\Delta G/4\pi = \delta(x) - 1$. A specific choice is

$$G(x) = \pi^{-1} \sum_{\substack{k \in Z^3 \\ k \neq 0}} |k|^{-2} \exp[2\pi i k \cdot x].$$

Theorem 2.16. ϕ , ρ and E satisfy

$$(i) E = (\gamma/10) \int_{\Gamma} \rho^{5/3} + (z/2) \lim_{x \rightarrow 0} \{ \phi(x) - z|x|^{-1} \}, \quad (2.25)$$

$$(ii) \phi(x) = zG(x) - \int_{\Gamma} G(x-y)\rho(y) + \psi_0 \quad (2.26a)$$

for some ψ_0 . Alternatively,

$$-\Delta\phi(x)/4\pi = \sum_{y \in Z^3} z\delta(x-y) - \rho(x), \quad (2.26b)$$

(iii) ϕ and ρ are real analytic on $R^3 \setminus Z^3$.

(iv) There is a unique pair ρ, ψ_0 that satisfies Eq. (2.26) with $\gamma\rho^{2/3} = \phi$ and $\int \rho = z$ (cf. Theorem 2.6).

Formula (2.25) may appear strange but it is obtained simply from the TF equation; an analogous formula also holds for a finite molecule.

Equation (2.26), together with $\gamma\rho^{2/3} = \phi$, is the periodic TF equation. ψ_0 is not a chemical potential. The chemical potential, $-\mu$, is zero because μ_{Λ} is zero for every finite system. If (2.26) is integrated over Γ we find, since $\int \rho = z$, that $\psi_0 = \int_{\Gamma} \phi =$ average electric potential. It might be thought that ψ_0 could be calculated in the same way that the Madelung potential is calculated: In each cubic cell there is (in the limit) a charge density

$z\delta(x) - \rho(x)$. Therefore if

$$\tilde{\phi}(x) \equiv \sum_{y \in Z^3} g(x-y),$$

with

$$g(x) = z|x|^{-1} - \int_{\Gamma} \rho(y)|x-y|^{-1} dy,$$

it might be expected that $\tilde{\phi} = \phi$. The correct statement is that $\phi(x) = \tilde{\phi}(x) + d$ and $d \neq 0$ in general. One can show that $\int_{\Gamma} \tilde{\phi} = 2\pi \int_{\Gamma} x^2 \rho(x) dx$ (see LS). The fact that $d \neq 0$, precludes having a simple expression for ψ_0 . Why is $d \neq 0$, i.e., why is $\phi \neq \tilde{\phi}$? The reason is that the charge density in the cell centered at $y \in Z^3$ is $z\delta(x-y) - \rho(x-y)$ only in the limit $\Lambda \rightarrow \infty$. For any finite Λ there are cells near the surface of Λ that do not yet have this charge distribution. Thus $d \neq 0$ essentially because of a neutral double layer of charge on the surface.

In LS asymptotic formulas as $z \rightarrow 0$ and ∞ are given for the various quantities.

Theorems 2.15 and 2.16 will not be proved here. Teller's lemma, which implies that $\phi_{\Lambda}(x)$ is monotone increasing in Λ , is used repeatedly. Apart from this, the analysis is reasonably straightforward.

G. The Thomas-Fermi theory of screening

Another interesting solid-state problem is to calculate the potential generated by one impurity nucleus, the other nuclei being smeared out into a uniform positive background (jellium model). If Λ is any bounded, measurable set in R^3 , and if $\rho_B = (\text{const}) > 0$ is the charge density of the positive background in Λ , and if the impurity nucleus has charge $z > 0$ and is located at 0, then the potential is

$$V_{\Lambda}(x) = z|x|^{-1} + \rho_B \int_{\Lambda} |x-y|^{-1} dy. \quad (2.27)$$

The TF energy functional, without the nuclear repulsion, and with $\gamma=1$, is

$$\mathcal{E}_{\Lambda}(\rho) = \frac{3}{5} \int \rho^{5/3} - \int V_{\Lambda} \rho + D(\rho, \rho). \quad (2.28)$$

The integrals are over R^3 , not Λ . Let $\rho_{\Lambda}(x)$ be the neutral minimizing ρ (so that $\int \rho_{\Lambda} = z + \rho_B |\Lambda|$).

Definition. A sequence of domains Λ in R^3 is said to *tend to infinity weakly* if every bounded subset of R^3 is eventually contained in Λ .

Remark. This is an extremely weak notion of $\Lambda \rightarrow \infty$.

It is intuitively clear that if $\Lambda \rightarrow \infty$ weakly and $z=0$ then $\rho_{\Lambda}(x) \rightarrow \rho_B$. For $z \neq 0$, $\rho_{\Lambda}(x) - \rho_B$ is expected to approach some function which looks like a Yukawa potential for large $|x|$. This is stated in many textbooks and is correct except for one thing: The coefficient of the Yukawa potential is *not* z but is some smaller number. In TF theory there is *over-screening* because of the nonlinearities.

Theorem 2.17. Let $\Lambda \rightarrow \infty$ weakly and $z=0$. Then $\phi_{\Lambda}(x) \rightarrow \rho_B^{2/3}$ uniformly on compacts in R^3 .

The theorem is another example of the effects of "surface charge." Since $\rho_{\Lambda} - \rho_B$ and $\phi_{\Lambda} = \rho_{\Lambda}^{2/3}$, the result is

natural. But it means that the average potential is not zero. If, on the other hand, the integrals in Eq. (2.28) are restricted to Λ then $\rho_\Lambda(x) = \rho_B$ for all Λ and $x \in \Lambda$, and $\phi_\Lambda(x) = 0$.

Theorem 2.18. *Let $\Lambda \rightarrow \infty$ weakly and $z > 0$. Let*

$$f(x) = \lim_{\Lambda \rightarrow \infty} \phi_\Lambda(x) - \rho_B^{2/3}$$

and

$$g(x) = \lim_{\Lambda \rightarrow \infty} \rho_\Lambda(x) - \rho_B.$$

- (i) *these limits exist uniformly on compacts,*
- (ii) $g \in L^1 \cap L^{5/3}$,
- (iii) $0 \leq f(x) \leq \phi^{\text{atom}}(x)$,
- (iv) *f and g are strictly positive and real analytic away from $x=0$,*
- (v) *$f(x)$ is monotone increasing in z ,*
- (vi) *These limits satisfy the TF equation*

$$f(x) = z|x|^{-1} - \int |x-y|^{-1} g(y) dy, \quad (2.29)$$

$$[\rho_B^{2/3} + f(x)]^{3/2} - \rho_B = g(x), \quad (2.30)$$

$$\int g = z, \quad (2.31)$$

(vii) *Assuming only that $g \in L^1 \cap L^{5/3}$ and $f(x) \geq -\rho_B^{2/3}$ there is only one solution to Eqs. (2.29) and (2.30) [without assuming (2.31)].*

There is a scaling relation:

$$f(x; z) = \rho_B^{2/3} F(\rho_B^{1/6} |x|; \rho_B^{-1/2} z)$$

$$g(x; z) = \rho_B G(\rho_B^{1/6} |x|; \rho_B^{-1/2} z).$$

Let us write $F(r; z) = q(r; z)Y(r)$ where $Y(r) = (1/r) \times \exp\{- (6\pi)^{1/2} r\}$ is the Yukawa potential.

Theorem 2.19. (i) $q(r; z)$ is monotone decreasing in r and increasing in z ;

(ii) $q(0; z) = z$;

(iii) $Q(z) = \lim_{r \rightarrow \infty} q(r; z)$ exists. $0 < Q(z) < z$ and Q is monotone increasing. $\limsup_{z \rightarrow \infty} Q(z)(6z)^{-2/3} < 1$ with $b = 1.039$.

LS contains graphical plots of $Q(z)$ and $q(r; 53.7)$. An asymptotic formula for $Q(z)$ has not been given. In the linearized approximations found in textbooks, $Q(z) = z$, but we see that this is false.

H. The Firsov variational principle

The problem of minimizing $\mathcal{E}(\rho)$ is a convex minimization problem. It has a dual which we now explore. The advantage of the dual problem is that it gives a lower bound to E . The principle was first given and applied in (Firsov, 1957) in the neutral case ($\mu = 0$) and was first rigorously justified in that case by Benguria (1979). Here we shall also state and prove the principle for non-neutral systems; furthermore, in the neutral case our (and Benguria's) principle will contain a slight improvement over Firsov's.

The dual functional to be considered is

$$\begin{aligned} \mathfrak{F}_\mu(f) = & - (8\pi)^{-1} \int |\nabla f(x)|^2 dx \\ & - \frac{2}{5} \gamma^{-3/2} \int [V(x) - f(x) - \mu]^{5/2} dx + U, \end{aligned} \quad (2.32)$$

where μ is a real parameter. The domain of $\mathfrak{F}_\mu$ is $B = \{f | \nabla f \in L^2, |f(x)| < c|x|^{-1} \text{ for some } c < \infty \text{ and for } |x| > R \text{ for some } R\}$. V is assumed to go to zero at ∞ and is such that the TF problem has a minimum for that V , and the minimizing ρ (with $\int \rho \leq \lambda$) satisfies the TF equation (2.18) (for all λ). We define

$$E^F(\mu) = \sup \{ \mathfrak{F}_\mu(f) | f \in B \} \quad (2.33)$$

Remark. When $\mu = 0$, Firsov imposed the additional constraint $V \geq f$. This, as we shall see, is unnecessary provided $[]_*^{5/2}$ is used as in Eq. (2.32).

Theorem 2.20. *If $\mu < 0$ then $E^F(\mu) = -\infty$. If $\mu \geq 0$ then there is a unique maximizing f for $\mathfrak{F}_\mu$. This f is $f_\mu \equiv |x|^{-1} \rho_\mu$ where ρ_μ is the unique solution to Eq. (2.18). If $\lambda = \int \rho_\mu$ then (see remark below)*

$$E^F(\mu) = E(\lambda) + \mu \lambda. \quad (2.34)$$

Proof. Suppose $\mu < 0$. Since V and any $f \in B \rightarrow 0$ as $|x| \rightarrow \infty$, the second term in (2.32) is $-\infty$. Suppose $\mu \geq 0$. Let $\tilde{E}_\mu =$ right side of (2.34). Clearly $\mathfrak{F}_\mu(f_\mu) = \tilde{E}_\mu$ by the TF equation (2.18). $f - \mathfrak{F}_\mu(f)$ is strictly concave because $\int (\nabla f)^2$ is strictly convex. Thus there can be at most one maximizing f , and we therefore must show that if $f \neq f_\mu$ then $\mathfrak{F}_\mu(f) \leq \tilde{E}_\mu$. By Minkowski's inequality ($|ab| \leq 2|a|^{5/2}/5 + 3|b|^{5/3}/5$) we have

$$\begin{aligned} \left(\frac{2}{5}\right)(V - f - \mu)^{5/2} \geq & - \left(\frac{3}{5}\right)(V - f_\mu - \mu)^{5/2} \\ & + (V - f - \mu)(V - f_\mu - \mu)^{3/2}. \end{aligned}$$

But $(V - f - \mu)_+ \geq V - f - \mu$, so $\mathfrak{F}_\mu(f) \leq \tilde{E}_\mu + h(f)$ where

$$h(f) = - (8\pi)^{-1} \int (\nabla f)^2 + \int f \rho_\mu - D(\rho_\mu, \rho_\mu).$$

By standard methods (e.g., Fourier transforms), $h(f) \leq 0$. Furthermore, $h(f) = 0$ only for $f = f_\mu$, which shows once again that the maximizing f is uniquely f_μ . ■

It should be noted that $E^F(\mu)$ is the Legendre transform of $E(\lambda)$. Namely $\lambda \rightarrow E(\lambda)$ is convex and

$$E^F(\mu) = \inf_{\lambda \geq 0} [E(\lambda) + \mu \lambda], \quad \text{all } \mu \in \mathbb{R}. \quad (2.35)$$

This shows that $E^F(\mu)$ is concave in μ . On the other hand, Theorem 2.20 displays $E^F(\mu)$ as the supremum (not infimum) of a family of concave functions. Furthermore, since $E(\lambda)$ is convex and bounded it is its own double Legendre transform, viz.

$$E(\lambda) = \sup_{\mu} [E^F(\mu) - \mu \lambda]. \quad (2.36)$$

Theorem 2.21. *Fix $\lambda \geq 0$. Then [by Eq. (2.36)]*

$$\sup \{ \mathfrak{F}_\mu(f) - \mu \lambda | f \in B, \mu \in \mathbb{R} \} = E(\lambda). \quad (2.37)$$

Remark. In Theorem 2.20 we refer to the unique ρ_μ satisfying Eq. (2.18) for $\mu \geq 0$. This requires some explanation. If $V(x)$ is unbounded (e.g., point nuclei), then as μ goes from ∞ to 0, λ goes from 0 to λ_c and $\rho_\mu(\lambda)$ minimizes $\mathcal{E}$ on $\int \rho = \lambda$. If $\text{ess sup } V(x) = v < \infty$, then $\rho_\mu \equiv 0$ [and $E^F(\mu) = 0$] for $\infty > \mu \geq v$. In this range $\lambda(\mu) = 0$. Then, as μ goes from v to 0, λ goes from 0 to λ_c , and $\rho_{\mu(\lambda)}$ minimizes $\mathcal{E}$ on $\int \rho = \lambda$. (ess sup is defined in Theorem 3.12).

III. THE "NO-BINDING" AND RELATED POTENTIAL-THEORETIC THEOREMS

The no-binding theorem was discovered by Teller (1962) and is one of the most important facts about the TF and TFD theories of atoms and molecules. It "explained" the absence of binding found numerically by Sheldon (1955). That this crucial theorem was not proved until 1962—after 35 years of intensive study of TF theory—is remarkable. It can be considered to be a prime example of the fact that pure analysis can sometimes be superior to numerical studies.

While Teller's ideas were correct, his proof was questioned on grounds of rigor. Balázs (1967) found a different proof for the special case of the symmetric diatomic molecule. A rigorous transcription of Teller's ideas was given in LS. In any case, all proofs of the theorem rely heavily on the fact that the potential is Coulombic.

There are really two kinds of theorems. An example of the first kind is "Teller's lemma," which states that the potential increases when nuclear charge is added. The second, "Teller's theorem" is the no-binding Theorem 3.23. The second, but not the first, requires the nuclear repulsion U . If U is dropped then the theorem goes the other way. The proof of Teller's theorem given in LS is complicated in the non-neutral case, but recently Baxter (1980) found a much nicer proof—one which actually produces a variational ρ that lowers the energy for separated molecules. Baxter's proposition (proposition 3.24) will appear again in Lemma 7.22.

In this section we shall consider general V and assume that

$$\mathcal{E}(\rho) = \int j(\rho(x)) dx - \int V(x)\rho(x) dx + D(\rho, \rho), \quad (3.1)$$

where j is a C^1 convex function with $j(0) = j'(0) = 0$. Note that in this section (only) $\mathcal{E}(\rho)$ does not contain U . This is done partly for convenience, but mainly for the reason that since V is not necessarily Coulombic the definition of U would have no clear meaning.

The Euler-Lagrange equation for (3.1) and $\rho(x) \geq 0$ is (with $\phi_\rho = V - |x|^{-1} * \rho$):

$$\begin{aligned} \phi_\rho(x) - \mu &= j'(\rho(x)) \quad \text{a.e. when } \rho(x) > 0, \\ &\leq 0 \quad \text{a.e. when } \rho(x) = 0. \end{aligned} \quad (3.2)$$

Any solution to (3.2) is determined only almost everywhere (a.e.).

We could, in fact, allow more general j 's of the form $j(\rho, x)$ [and $\int j(\rho(x), x) dx$ in $\mathcal{E}$] with $j(\cdot, x)$ having the above properties for all x , but we shall not do so. An annoying case we must consider, however, is $j'(\rho) = 0$ for $0 < \rho < \rho_0$ and $j'(\rho) > 0$ for $\rho > \rho_0$. This is discussed in some detail in Sec. III.C and is needed for TFD theory (Sec. VI). If $j'(\rho) > 0$, all $\rho > 0$, as it is in TF theory with $j'(\rho) = \gamma\rho^{2/3}$, then Eq. (3.2) can be written as

$$(\phi_\rho(x) - \mu)_+ \equiv \max[\phi_\rho(x) - \mu, 0] = j'(\rho(x)), \quad (3.2')$$

but otherwise (3.2) is stronger than (3.2').

One aim of this section is to study solutions of Eq. (3.2) without considering whether or not (3.2) truly comes from minimizing (3.1) or assuming uniqueness.

Definition. $\mathfrak{e} = \{\rho \mid \rho(x) \geq 0, \rho \in L^1, \text{ and } \int \rho(y)|x-y|^{-1} dy$

is a bounded, continuous function which goes to zero as $x \rightarrow \infty\}$.

We shall be concerned only with solutions to (3.2) in $\mathfrak{e}$.

The following lemma (LS, II.25) is useful, in the cases of interest, to guarantee that $\rho \in \mathfrak{e}$.

Lemma 3.1. *If $f \in L^p, g \in L^{p'}$, $1/p + 1/p' = 1$, $p, p' > 1$ then $f * g$ is a bounded, continuous function which goes to zero as x goes to infinity. In particular, if $\rho \in L^{3/2+\epsilon} \cap L^1$ then $\rho \in L^{3/2+\epsilon} \cap L^{3/2-\epsilon}$. Since $|x|^{-1} \in L^{3+\epsilon} + L^{3-\epsilon}$, $\rho \in \mathfrak{e}$.*

It will always be assumed that $V(x) \rightarrow 0$ as $|x| \rightarrow \infty$ (this always means uniformly with respect to direction). Hence μ cannot be negative in Eq. (3.2), for otherwise $\rho \notin L^1$.

A. Some variational principles and Teller's lemma

At first it will *not* be assumed that V is Coulombic.

Theorem 3.2. *Fix $\lambda > 0$ and suppose that $\rho_\lambda, \mu_\lambda$ satisfy Eq. (3.2) with $\int \rho_\lambda = \lambda$. Let $\phi_\lambda = \phi_{\rho_\lambda}$ and assume that $\rho_\lambda \in \mathfrak{e}$. Then, for all x ,*

- (a) $\phi_\lambda(x) - \mu_\lambda = \sup_{\rho} \left\{ \phi_\rho(x) - \mu \mid \phi_\rho(y) - \mu \leq j'(\rho(y)) \text{ a.e. } y, \int \rho \leq \lambda, \rho \in \mathfrak{e} \right\},$
- (b) $\phi_\lambda(x) - \mu_\lambda = \inf_{\rho} \left\{ \phi_\rho(x) - \mu \mid \phi_\rho(y) - \mu \geq j'(\rho(y)) \text{ a.e. } y \text{ when } \rho(y) > 0, \int \rho \geq \lambda, \rho \in \mathfrak{e} \right\}$
- (c) $\phi_\lambda(x) = \sup_{\rho} \left\{ \phi_\rho(x) \mid \phi_\rho(y) - \mu_\lambda \leq j'(\rho(y)), \text{ a.e. } y, \rho \in \mathfrak{e} \right\}$
- (d) $\phi_\lambda(x) = \inf_{\rho} \left\{ \phi_\rho(x) \mid \phi_\rho(y) - \mu_\lambda \geq j'(\rho(y)), \text{ a.e. } y \text{ when } \rho(y) > 0, \rho \in \mathfrak{e} \right\}.$

Furthermore, in (a) [resp. (b)] there is no ρ satisfying the conditions on the right when $\mu < \mu_\lambda$ [resp. $\mu > \mu_\lambda$]. Note that in (a) and (b) μ is arbitrary (including $\mu < 0$) and ρ is constrained, while in (c) and (d) the opposite is true [except, of course, $\rho(x) \geq 0$].

In the following, a statement such as $\Delta\phi/4\pi = \rho$ is always meant in the distributional sense. We shall need Lemma II.26 from LS.

Lemma 3.3. *Let $\rho_1, \rho_2 \in L^1$ with $\rho_1(x) \geq 0$ and $\psi_i = |x|^{-1} * \rho_i$. If $\psi_1(x) \geq \psi_2(x)$, all x , then $\int \rho_1 \geq \int \rho_2$.*

Proof. Suppose $\int \rho_2 - \rho_1 = 4\epsilon > 0$. There exists a ball B , of radius R , such that $\int \rho_1(1 - \Theta) < \epsilon$ and $\int \rho_2(1 - \Theta) < \epsilon$, where $\Theta(x) = 1$ for $x \leq R$ and zero otherwise. Compute the spherical average of $\psi_2 - \psi_1$ on the sphere of radius R ; it cannot be positive. The contribution from inside B is, by Newton's theorem, $R^{-1} \int (\rho_2 - \rho_1)\Theta > 2\epsilon/R$. The contribution from outside B is at least $-R^{-1} \int \rho_1(1 - \Theta) > -\epsilon/R$. Adding these gives a contradiction. ■

Remark. Even if $\psi_1(x) > \psi_2(x)$ for all x , we cannot conclude that $\int \rho_1 > \int \rho_2$.

Proof of Theorem 3.2. (a) will be proved here; (b), (c), and (d) follow similarly. Since ρ_λ gives equality, $\phi_\lambda - \mu_\lambda \leq \sup \{ \dots \}$. We have to show that if $(\phi_\rho - \mu)_+ \leq j'(\rho)$

a.e. and if $\int \rho \leq \lambda$ then

- (i) $\phi_\lambda(x) - \mu_\lambda \geq \phi_\rho(x) - \mu$,
- (ii) $\mu \geq \mu_\lambda$.

First suppose $\mu \geq \mu_\lambda$ and let $\psi(x) = \phi_\rho(x) - \phi_\lambda(x) + \mu_\lambda - \mu$. Let $B = \{x | \psi(x) > 0\}$. B is open since ψ is continuous. As a distribution $-(4\pi)^{-1} \Delta \psi(x) = \rho_\lambda(x) - \rho(x) \leq 0$ a.e. on B since $j'(\rho) \geq \phi_\rho - \mu$, $j'(\rho_\lambda) = \phi_\lambda - \mu_\lambda$ when $\rho_\lambda > 0$ and j' is nondecreasing. Hence ψ is subharmonic on B and takes its maximum on ∂B , the boundary of B , or at ∞ . $\psi = 0$ on ∂B . At ∞ , $\psi = \mu_\lambda - \mu \leq 0$. Hence B is empty and (i) is proved. Suppose now that $\mu_\lambda - \mu = \delta > 0$. Then $j'(\rho) \geq \phi_\rho - \mu > \phi_\rho - \mu_\lambda$ and, by the previous proof (applied to $\mu = \mu_\lambda$), $\phi_\lambda(x) \geq \phi_\rho(x)$. By Lemma 3.3, $\int \rho \geq \int \rho_\lambda = \lambda$. Hence $\int \rho = \lambda$. At this point there are two possible strategies.

(i) If we assume that $\mathcal{G}(\rho)$ has a minimum that satisfies Eq. (3.2) for all $\mu \geq 0$, then we can use the fact [which follows from the strict convexity of $\mathcal{G}(\rho)$] that μ_λ is a continuous decreasing function of λ . Then there exists $\gamma > \lambda$ with $\mu_\gamma > \mu_\lambda > \mu$. Since $j'(\rho) \geq \phi_\rho - \mu_\gamma$, $\int \rho = \gamma$ by what we just proved. But this is a contradiction.

(ii) There is a purely potential theoretic argument without invoking Eq. (3.1). There is a (not necessarily unique) f which satisfies $j'(f(x)) = [\phi_\lambda(x) - \mu/2 - \mu_\lambda/2]$ and $f(x) = 0$ when $|x| = 0$. Hence $f(x) < \rho(x)$ a.e. when $\rho(x) > 0$, and $f(x) = 0$ when $\rho(x) = 0$. Thus $f \in \mathcal{C}$. Since $\int \rho = \lambda > 0$, $\int f < \lambda$. Let $g = (1 - \varepsilon)\rho + \varepsilon f$, $0 < \varepsilon < 1$. Since $|x|^{-1} * \rho$ (and hence $|x|^{-1} * f$) are bounded, $\phi_\lambda(x) \leq \phi_g(x) + \varepsilon C$ for some constant C . Choose $\varepsilon > 0$ so that $\varepsilon C < \delta/2$. Then $j'(g) \geq j'(f) \geq \phi_g - \mu_\lambda$. Since $\int g < \lambda$, g satisfies the condition in (a) with $\mu = \mu_\lambda$ but, as we have seen, this implies $\int g = \lambda$. ■

Teller's lemma (Theorem 3.4) is closely related to Theorem 3.2.

Definition. We say $V \in \mathcal{D}$ if $V \neq 0$ and V is superharmonic, vanishing at ∞ (and hence $V > 0$). Moreover, the set $\{x | V(x) = \infty\} = S_V$ (called the *singularities* of V) is closed, V is continuous on the complement of S_V , and $V(x) \rightarrow \infty$ as $x \rightarrow S_V$.

Theorem 3.4. Suppose V is replaced by $V' = V + W$ with $W \in \mathcal{D}$. (In the case of interest $W = z|x - R|^{-1}$, which means that we add, or increase, a nuclear charge.) Suppose that for some common μ there are solutions to Eq. (3.2) $0 \leq \rho$, $\rho' \in \mathcal{C}$ with V and with V' . Then $\phi'(x) \geq \phi(x)$ all x and, if j' is strictly monotone or if $\phi - \phi' \in H^2$ (i.e., $\phi - \phi'$ and its first two derivatives are in L^2) away from S_W , then $\rho'(x) \geq \rho(x)$, a.e.

Proof. Let $\psi = \phi' - \phi$ and $B = \{x | \psi(x) < 0\}$. Clearly $B \cap S_W = \emptyset$ so B is open. As a distribution, $\Delta \psi/4\pi \leq \rho' - \rho \leq 0$, so ψ is superharmonic on B . Thus B is empty and $\phi' \geq \phi$. The proof that $\rho' \geq \rho$ is trickier. If j' is strictly monotone it is obvious. Otherwise it can be shown (see Benguria, 1979) that for suitable V , W , and j' , $\psi \in H^2$ away from S_W . Assuming $\psi \in H^2$, if $\rho'(x) < \rho(x)$ then $x \in C = \{x | \psi(x) = 0\}$. On C , $\Delta \psi = 0$, a.e. (see Benguria, 1979, Theorems 2.19 and 3.3). Let $D = C \cap \{x | \rho'(x) < \rho(x)\}$. On D , $0 = \Delta \psi/4\pi \leq \rho' - \rho < 0$ a.e., so D has zero measure. ■

Remark. If $j'(s)$ is strictly monotone and $W \neq 0$ then $\phi'(x) > \phi(x)$ for all $x \notin S_W$.

A similar proof yields

Theorem 3.5. If $V \in \mathcal{D}$ then $\phi_\lambda(x) \geq 0$. Consequently if $V(x) = \int dM(y) |x - y|^{-1}$, $dM \geq 0$, and $\int dM = Z$, then there is no solution if $\lambda > Z$ because then $\phi_\lambda(x) < 0$ for some large x . Cf. Theorem 6.7.

There are many easy, but important corollaries of Theorem 3.2. We stress that V need not be Coulombic; the important ingredient is that the electron-electron repulsion is Coulombic.

Definition. j' is said to be *subadditive* if $j'(\rho_1 + \rho_2) \leq j'(\rho_1) + j'(\rho_2)$. j' is subadditive in the TF case.

Corollary 3.6. Suppose $V = V_1 + V_2$ and μ is fixed. Let ϕ_1, ϕ_2 be solutions to Eq. (3.2) for this μ with V_1, V_2 , respectively. Suppose $\phi_i \geq 0$ (e.g., $V_i \in \mathcal{D}$) and suppose j' is subadditive. Then $\phi(x) \leq \phi_1(x) + \phi_2(x)$, all x .

Proof. Use Theorem 3.2 (d) with $\rho_1 + \rho_2$ on the right side. ■

Corollary 3.7. Let $\lambda > 0$. There can be at most one pair ρ, μ satisfying Eq. (3.2) with $\rho \in \mathcal{C}$ (in particular for $\rho \in L^{5/3} \cap L^1$) and $\int \rho = \lambda$.

Proof. If ρ_1, ρ_2 are two solutions, use Theorem 3.2(a) twice with ρ_1 and ρ_2 to deduce $\phi_1 - \mu_1 = \phi_2 - \mu_2$. This implies $\mu_1 = \mu_2$ and hence $\phi_1 = \phi_2$. But then $0 = \Delta(\phi_1 - \phi_2) = 4\pi(\rho_2 - \rho_1)$. ■

This uniqueness result was proved earlier, Theorem 2.6, using the strict convexity of $\mathcal{G}(\rho)$.

Corollary 3.8. If $0 < \lambda' < \lambda$ then

- (i) $\phi_{\lambda'} \geq \phi_\lambda$
- (ii) $\mu_{\lambda'} \geq \mu_\lambda$
- (iii) $\phi_{\lambda'} - \mu_{\lambda'} \leq \phi_\lambda - \mu_\lambda$.

Proof. For (iii) use Theorem 3.2(b) with $\rho_\lambda, \mu_\lambda$ as trial function for the λ' problem. (iii) $\Rightarrow$ (ii). For (i) use (c) with ρ_λ as variational function for the λ' problem. ■

Corollary 3.9. Suppose ρ_1, μ and ρ_2, μ (same μ) are two solutions to Eq. (3.2) with $\int \rho_i = 0$. Then $\phi_1 = \phi_2$ and $\rho_1 = \rho_2$ a.e. Therefore, by Corollary 3.8, whenever $\lambda_2 > \lambda_1$ then $\mu_2 < \mu_1$ (i.e., $\mu_2 = \mu_1$ cannot occur).

Proof. Using Theorem 3.2(d), $\phi_1 = \phi_2$. Then

$$0 = \Delta(\phi_1 - \phi_2)/4\pi = \rho_1 - \rho_2 \text{ a.e.} \quad \blacksquare$$

Corollary 3.10. Suppose $j'_1(\rho) \leq j'_2(\rho)$, all ρ . Let $\rho_\lambda^1, \mu_\lambda^1$ and $\rho_\lambda^2, \mu_\lambda^2$ be corresponding solutions to Eq. (3.2) with fixed λ , and ρ^1, ρ^2 solutions with fixed μ . $\phi_\lambda^i(x)$ are the corresponding potentials. Then

- (i) $\phi_\lambda^1 - \mu_\lambda^1 \leq \phi_\lambda^2 - \mu_\lambda^2$,
- (ii) $\mu_\lambda^1 \geq \mu_\lambda^2$,
- (iii) $\phi^1 \leq \phi^2$.

Proof. For (i) use Theorem 3.2(a) with $\rho_\lambda^1, \mu_\lambda^1$ as trial function for the 2 problem, (i) $\Rightarrow$ (ii). For (iii) use (d) with ρ_2 as trial function for the 1 problem. ■

Lemma 3.11. When $\mu > 0$, ρ has compact support.

Remark. As will be seen in Sec. VI, ρ has compact support in TFD theory even when $\mu = 0$. See Theorem 6.6.

Among the most important consequences of Theorem

3.2 are the variational principles for the chemical potential [LS].

Theorem 3.12. Define the functionals

$$T(\rho) = \text{ess sup} \{ \phi_\rho(x) - j'(\rho(x)) \}, \quad (3.3)$$

$$S(\rho) = \text{ess inf}_{x: \rho(x) > 0} \{ \phi_\rho(x) - j'(\rho(x)) \}.$$

(ess sup means supremum modulo sets of measure zero).

Then, whenever there is a solution to (3.2) with $\int \rho = \lambda > 0$,

$$\mu_\lambda = \inf \left\{ T(\rho) \mid \rho \in \mathcal{C}, \int \rho \leq \lambda \right\} \quad (3.4)$$

$$\mu_\lambda = \sup \left\{ S(\rho) \mid \rho \in \mathcal{C}, \int \rho \geq \lambda \right\}. \quad (3.5)$$

Corollary 3.13. If $j'(p)$ is concave (as in TF theory with $j' = \rho^{2/3}$) then μ_λ and $\mu_\lambda - \phi_\lambda(x)$, for each x , are jointly convex functions of V and λ .

Corollary 3.14. If λ is fixed and $V_1(x) \geq V_2(x)$, all x , then $\mu_1(1) \geq \mu_\lambda(2)$.

By Corollaries 3.9 and 3.14 we know that increasing V increases μ while increasing λ decreases μ . What happens if V and λ are both increased, in particular if we scale up the size of a molecule by $V \rightarrow \alpha V$, $\lambda \rightarrow \alpha \lambda$? A partial answer is given by the following two corollaries.

Corollary 3.15. Let $V_1, V_2 \in \mathcal{D}$ and $V = V_1 + V_2$. Assume j' is subadditive and suppose Eq. (3.2) has solutions to the three problems (V_1, λ_1) , (V_2, λ_2) , (V, λ) with $\lambda = \lambda_1 + \lambda_2$. Then $\mu \geq \min(\mu_1, \mu_2)$.

Proof. In general, if $W \in \mathcal{D}$ and ρ is a solution to (3.2) with W then $\phi_\rho(x) - j'(\rho(x)) = \mu$ a.e. if $\rho(x) > 0$ and ≥ 0 a.e. if $\rho(x) = 0$ (Theorem 3.5). From this remark it follows that $S_V(\rho_1 + \rho_2) \geq \min(\mu_1, \mu_2)$. [Here, $S_V(\rho)$ refers to Eq. (3.3) with V .] ■

Corollary 3.16. Let $\alpha > 1$ and suppose Eq. (3.2) has solutions with (V, λ, μ) and $(\alpha V, \alpha \lambda, \mu(\alpha))$. Assume j' satisfies $j'(\alpha t) \leq \alpha j'(t)$, all t (this holds in TF theory). Then $\mu(\alpha) \geq \alpha \mu$.

Proof. If ρ is the solution to (V, λ, μ) then $S_V(\rho) = \mu$. But $S_{\alpha V}(\alpha \rho) \geq \alpha S_V(\rho)$. ■

Corollary 3.17. Suppose there is a solution to Eq. (3.2) for all $\lambda \in (a, b)$ with $a < b$. Then μ_λ is continuous on this interval.

Proof. Let $\lambda_2 = \lambda_1 + \varepsilon$. By Corollary 3.9, $\mu_1 > \mu_2$. Let $\rho = \rho_1 + \varepsilon \chi$ with $\int \chi = 1$, $0 \leq \chi(x) \leq b$ for some b , $\chi(x) = 0$ if $\rho_1(x) = 0$, and $\chi(x) = 0$ if $\rho_1(x) > a$ for some a . Then $\chi \in \mathcal{C}$. Since j' is continuous, $S(\rho) \geq \mu_1 - Q(\varepsilon)$ where $Q(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. ■

Theorem 3.18. Let $V \in \mathcal{D}$, $V(x) = \int dM(y) |y - x|^{-1}$, $dM \geq 0$, $\int dM = Z > 0$. Suppose that for large t , $j'(t) > ct^{(1/2)+\varepsilon}$, with $c, \varepsilon > 0$. By a simple modification of the method of Theorems 2.4, 2.5, and 2.6, $\mathcal{S}(\rho)$ has a unique minimum on the set $\mathcal{C}$ with $\int \rho \leq \lambda$. This ρ satisfies Eq. (3.2) and $\int \rho = \lambda$ if $\lambda \leq \lambda_c$, whereas $\int \rho = \lambda_c$ if $\lambda > \lambda_c$. Now assume, in addition, that $j'(t) < dt^{\varepsilon+1/3}$, $\varepsilon > 0$, for small t (this is true in all cases of interest). Then $\lambda_c = Z$.

Proof. If $\lambda > Z$ there is no solution by Theorem 3.5.

Now suppose $\mu = 0$; we claim $\lambda \geq Z$, and hence that $\lambda = Z$. If so, we are done because $\mathcal{S}(\rho)$ has an absolute minimum. This minimum corresponds to $\mu = 0$ and has $\lambda = \lambda_c$; but $\mu = 0$ implies $\lambda = Z$. Now, to prove that $\lambda \geq Z$, let ϕ be the solution. If $\lambda = Z - 3\delta$, let χ be the characteristic function of a ball centered at the origin such that $\int \chi dM > Z - \delta$. Then

$$\phi(x) > \psi(x) = \int [\chi(y)dM(y) - \rho(y)dy] |x - y|^{-1}.$$

For $|x| > \text{some } R$, $\psi(x) < 2Z|x|^{-1}$. Also $[\psi(x)] =$ (spherical average of ψ) $> 2\delta|x|^{-1}$ for $x > R$. For a given $|x| = r > R$ let $\Omega_r(r)$ be the proportion of the sphere of radius r such that $2Z > r\psi(x) > \delta$, and let $\Omega_r(r)$ be the complement. Then $2\delta < r[\psi(x)] < 2Z\Omega_r + \delta\Omega_r = \delta + (2Z - \delta)\Omega_r$. Thus $\Omega_r(r) > \delta/(2Z - \delta)$ for all r . On Ω_r , $\rho^{(1/3)+\varepsilon} > \delta|x|^{-1}$ for large $|x|$, and therefore $\rho \notin L^1$ if $\delta > 0$. ■

Brezis and Benilan (Brezis, 1978, 1980) have generalized this. Even if $j(\rho) \sim \rho^{(4/3)-\varepsilon}$ for large ρ there is a solution to Eq. (3.2) if $\lambda \leq Z$, and no solution otherwise. This is noteworthy, since if $j(\rho) \sim \rho^a$ for large ρ with $a \leq \frac{2}{3}$ then $\mathcal{S}(\rho)$ has no lower bound for point nuclei. There are similar results for other potentials, V , in LS, Theorem II.18.

There is also an "energetic," as distinct from potential theoretic, reason that there is no solution if $\lambda > Z$. A solution to Eq. (3.2) implies a minimum for the functional $\mathcal{S}(\rho)$, by strict convexity. If $\lambda = \int \rho > Z$ then ϕ is negative in some set A of positive measure. Then it is easy to see that if ρ is decreased slightly in A to $\tilde{\rho}$, then $\mathcal{S}(\tilde{\rho}) < \mathcal{S}(\rho)$. But $\tilde{\rho} < \lambda$ and $E(\lambda)$ is nonincreasing.

In the variational principle, Theorem 3.12, ρ_λ gives equality, i.e., $T(\rho_\lambda) = S(\rho_\lambda) = \mu_\lambda$. Is this the only ρ with this property? If $\lambda > Z$ there are many ρ 's with $T(\rho) = 0$ and no ρ with $S(\rho) = 0$ (cf. LS). In Brezis, 1980, Sec. 4, it is shown that if j' is concave (as in TF theory) and V has suitable properties (satisfied for $V \in \mathcal{D}$) then when $\lambda < Z$ only ρ_λ satisfies either $T(\rho) = \mu_\lambda$ or $S(\rho) = \mu_\lambda$. If $\lambda = Z$ this uniqueness is lost in general!

Asymptotics of the chemical potential.

Theorem 3.12 can be used to obtain bounds on μ_λ . In the TF case with point nuclei, the asymptotic formula

$$\mu_\lambda \sim \gamma^{-1}(\pi^2 \sum z_j^3 / 4\lambda)^{2/3} \quad (3.6)$$

holds for λ small (LS, Theorem II.31). For λ near Z LS (Theorems IV.11, 12) find upper and lower bounds for μ_λ of the form $\alpha_\lambda(Z - \lambda)^{4/3}$ with $Z = \sum z_j$. Brezis and Benilan (unpublished) have shown that

$$\alpha = \lim_{\lambda \nearrow Z} \mu_\lambda(Z - \lambda)^{-4/3} \text{ exists} \quad (3.7)$$

and is given by solving some differential equation. α is independent of the number of nuclei and their individual coordinates and charges!

Equation (3.7) implies that there is a well defined ionization potential I in TF theory (although it probably has nothing to do with the true Schrödinger ionization energy). First observe that if we start with $\sum z_j = 1$ and then replace z_j by Zz_j , R_j by $Z^{-1/3}R_j$, and λ by $Z\lambda$, then by scaling Eq. (2.24),

$$\mu_{Z\lambda} = Z^{4/3}\mu_\lambda. \quad (3.8)$$

Therefore, by Eq. (3.7), if we let $\lambda = Z - \varepsilon$ with $\varepsilon > 0$ fixed, and let $Z \rightarrow \infty$, then

$$\lim_{Z \rightarrow \infty} \mu_{Z-\varepsilon} = \alpha \varepsilon^{4/3}. \quad (3.9)$$

The ionization potential is defined to be

$$I = E(\lambda = Z - 1) - E(\lambda = Z). \quad (3.10)$$

By integrating (3.9), and appealing to dominated convergence,

$$I \rightarrow 3\alpha/7 \text{ as } Z \rightarrow \infty. \quad (3.11)$$

Another implication of Eq. (3.7) is that an ionized atom has a well defined radius as $Z \rightarrow \infty$. This question was raised by Dyson. Suppose $V(x) = Z|x|^{-1}$ and $\lambda = Z - \varepsilon$. The density ρ will have support in a ball of radius $R(Z, \varepsilon)$. At $|x| = R$, $\phi(x) = \mu$. But since ρ is spherically symmetric, $R\phi(x) = Z - \lambda = \varepsilon$ by Newton's theorem. Thus the atomic radius satisfies

$$R = \varepsilon/\mu \text{ for all atoms} \quad (3.12)$$

and, by Eq. (3.9),

$$\lim_{Z \rightarrow \infty} R(Z, \varepsilon) = (\alpha \varepsilon^{1/3})^{-1}. \quad (3.13)$$

There are other ways in which TF theory yields a well defined atomic radius. See Sec. V.C (6).

B. The case of flat j' (TFD)

In TFD theory, as will be seen in Sec. VI, we have to consider

$$j'(\rho) = 0, \quad 0 \leq \rho \leq \rho_0 = (5C_e/8\gamma)^3 \\ = \gamma \rho^{2/3} - C_e \rho^{1/3} + 15C_e^2/4^3 \gamma, \quad \rho_0 \leq \rho. \quad (3.14)$$

j' satisfies all necessary conditions. It is neither concave nor subadditive, however. Let us consider V of the form

$$V(x) = \int dm(y) |x - y|^{-1}, \quad (3.15)$$

with m being a measure that is not necessarily positive. In the primary case of interest, $dm(x) = \sum z_j \delta(x - R_j)$.

The question we address here (and which will be important in Sec. VI) is this: Does $\rho(x)$ [the solution to Eq. (3.2)] take values in $(0, \rho_0)$? It may or may not, depending on m and λ .

Example. Suppose $dm(x) = g(x)dx$ with $g(x) \in (0, \rho_0)$ and $\int g = Z < \infty$. Then $\rho(x) = g(x)$ satisfies Eq. (3.2) with $\lambda = Z$, and thus $\rho(x) \in (0, \rho_0)$. This ρ also clearly minimizes $\mathcal{E}(\rho)$ in Eq. (3.1).

Nevertheless, in some circumstances $\rho \notin (0, \rho_0)$.

Theorem 3.19. Suppose $j'(\rho) = \alpha = \text{constant}$ for $\rho \in F \equiv (\rho_1, \rho_0]$ with $0 \leq \rho_1 \leq \rho_0 < \infty$, and $j'(\rho) > c\rho^{(1/2)+\varepsilon}$ for large ρ . Let V be given by Eq. (3.15) and let A be a bounded open set such that as distributions on A either $\rho_0 dx < dm < (\rho_0 + \text{const})dx$ or $dm < \rho_1 dx$. Let $\rho \in \mathcal{C}$ satisfy Eq. (3.2). Then $\rho(x) \notin F$ a.e. (with respect to Lebesgue measure) on A .

Proof. Cf Benguria, 1979, Lemmas 2.19, 3.2. First, it can be shown that $\phi_\rho \in H^2(A)$ (Sobolev space). Let $B = \{x | \rho(x) \in F\} \cap A$. On B , $\phi_\rho - \mu = \alpha$ and since $\phi_\rho \in H^2(A)$,

$\Delta \phi_\rho = 0$ a.e. on B . But $\Delta \phi_\rho / 4\pi = \rho - dm/dx$. ■

Remark. Since a solution to Eq. (3.2) is determined only a.e., $\rho(x)$ can be chosen $\notin F$ for all $x \in A$.

Corollary 3.20. Consider the TFD problem (3.14) with $V(x) = \sum z_j |x - R_j|^{-1}$. Then any solution to (3.2) can be modified on a set of measure zero so that $\rho(x) \notin (0, \rho_0]$ for all x .

C. No-binding theorems

Henceforth it will be assumed, as in Theorem 3.18, that j is such that Eq. (3.1) has a minimum for $\lambda \leq \lambda_c$ which satisfies Eq. (3.2). We shall be interested in comparing three (nonzero) potentials, V_1 , V_2 , and $V_{12} = V_1 + V_2$ with $V_i \in \mathcal{D}$. At first we shall consider what happens when the repulsion U is absent. As usual we define $e_a(\lambda) \equiv \inf \mathcal{E}_a(\rho)$ with $\lambda = \int \rho$ and $\mathcal{E}_a$ having V_a . There is no U term in $\mathcal{E}_a$, Eq. (3.1). Define

$$\Delta e(\lambda) = e_{12}(\lambda) - \min_{\lambda_1 + \lambda_2 = \lambda} e_1(\lambda_1) + e_2(\lambda_2). \quad (3.16)$$

Definition. If $\Delta e < 0$ (resp. ≥ 0) we say that in the absence of the repulsion U there is binding (resp. no binding).

Theorem 3.21. Suppose j satisfies

$$j(a+b) \leq j(a) + j(b) + aj'(b) + bj'(a), \quad a, b \geq 0. \quad (3.17)$$

[If j' is subadditive then Eq. (3.17) is satisfied. $j(t) = t^{5/3}$ satisfies (3.17).] Then $\Delta e < 0$.

Proof. For $i = 1, 2$ let λ_i minimize in Eq. (3.16) and let ρ_i be the minimizing ρ for $\mathcal{E}_i$ with $\int \rho_i \leq \lambda_i$. Recall $e_a(\lambda)$ is monotone nonincreasing. Let $\rho \equiv \rho_1 + \rho_2$ be a trial function for e_{12} in $\mathcal{E}_{12}$ and use the variational equations (3.2) for ρ_i and the fact that $\phi_i(x) \geq 0$. ■

Remark. The condition (3.17) is satisfied in TF theory but not in TFD theory.

Theorem 3.21 says we can [and do, if j satisfies Eq. (3.17)] have binding if the repulsion U is absent. The no-binding theorem, which we turn to now, relies on the addition of U which, by itself without $\mathcal{E}$, obviously has the no-binding property.

Proposition 3.22. If j is convex and $j(0) = 0$, then j has the superadditivity property: $j(a+b) \geq j(a) + j(b)$. If j' is strictly monotone, then the foregoing inequality is strict when $a, b \neq 0$.

Note. We assumed that j is convex in all cases. Therefore Theorem 3.23 holds in all cases.

Definition. Let

$$V_i = \frac{1}{|x|} * m_i \quad (m_i \text{ a measure})$$

be in $\mathcal{D}$. Then

$$D(m_1, m_2) \equiv \frac{1}{2} \int dm_1(x) dm_2(y) |x - y|^{-1}.$$

Theorem 3.23 (no binding). Let m_i , $i = 1, 2$ be non-negative measures of finite mass $z_i > 0$ and $V_i \in \mathcal{D}$. Then

$$\Delta E(\lambda) \equiv \Delta e(\lambda) + 2D(m_1, m_2) \geq 0. \quad (3.18)$$

If j is strictly superadditive then > 0 holds.

Remarks. Obviously $\Delta E(\lambda)$ is the energy difference when the repulsion U is included. *Binding never occurs.* In particular, if

$$V_1 = \sum_{j=1}^n z_j |x - R_j|, \quad V_2 = \sum_{j=n+1}^k z_j |x - R_j|^{-1}$$

then

$$m_1 = \sum_{j=1}^n z_j \delta(x - R_j), \quad m_2 = \sum_{j=n+1}^k z_j \delta(x - R_j).$$

In TFD theory j is not strictly superadditive. As we shall see in Sec. IV.C, it is possible to have a neutral diatomic molecule for which equality holds in Eq. (3.18).

Proof. We give two proofs. The LS proof in the neutral case $\lambda = z_1 + z_2$ is the following: Clearly $\lambda_1 = z_1$, $\lambda_2 = z_2$, $\mu_1 = \mu_2 = \mu_{12} = 0$. Consider $m_1 \rightarrow \alpha m_1$, $\lambda_1 \rightarrow \alpha z_1$, $0 \leq \alpha \leq 1$. By Theorem 2.13 we have

$$\partial e_f / \partial \alpha = - \int V_1 \rho_1$$

$$\partial e_{12} / \partial \alpha = - \int V_1 \rho_{12}.$$

Thus

$$\partial(e_{12} - e_1 + 2D(\alpha m_1, m_2)) / \partial \alpha = \int dm_1(x) [\phi_{12}(x) - \phi_1(x)].$$

But $\phi_{12}(x) \geq \phi_1(x)$, all x (Theorem 3.4 and following remark). When $\alpha = 0$, $\Delta E = 0$, so this proves the theorem. In the non-neutral case the $\mu_a \neq 0$ and it is necessary to take into account the change of μ_a with α . This is complicated (see LS).

The second proof is due to Baxter (1980). For any ρ_{12} with $\int \rho_{12} = \lambda$ we can, by Prop. 3.24, find g , $0 \leq g(x) \leq \rho_{12}(x)$, and $h(x) \equiv \rho_{12}(x) - g(x)$ such that $\psi_g(x) = \psi_{m_1}(x) = V_1(x)$ a.e. when $h(x) > 0$, and $\psi_g(x) \leq V_1(x)$ a.e. when $h(x) = 0$.

Let $a = \int g$, $b = \int h$. Then

$$\begin{aligned} \min \{ e_1(\lambda_1) + e_2(\lambda_2) \mid \lambda_1 + \lambda_2 = \lambda \} &\leq e_1(a) + e_2(b) \leq \mathcal{E}_1(g) + \mathcal{E}_2(h) \\ &\leq \mathcal{E}_{12}(\rho_{12}) + 2D(m_1, m_2) + \int h(V_1 - \psi_g) dx - \int (V_1 - \psi_g) dm_2 \\ &\leq e_{12}(\lambda) + 2D(m_1, m_2). \end{aligned} \quad (3.19)$$

The third inequality uses the superadditivity of j . If j' is strictly monotone this superadditivity is strict (and so is the final inequality) provided $g \neq \rho_{12}$. If $g = \rho_{12}$ a.e. then $\psi_{\rho_{12}} \leq V_1$ and hence $\lambda \leq z_1$ must hold. Choose $\lambda_1 = \lambda$, $\lambda_2 = 0$ and note that $e_1(\lambda) < \mathcal{E}_1(\rho_{12})$ because ρ_{12} does not satisfy Eq. (3.2) since $V_2 \neq 0$. Equation (3.19) then gives strict inequality. ■

Proposition 3.24 (Baxter, 1980). Let $V \in \mathcal{D}$ and let $\rho(x) \geq 0$ be a given function with $|x|^{-1} * \rho \in \mathcal{D}$. Assume $\rho \in L^p$ for some $p > \frac{3}{2}$ and $D(\rho, \rho) < \infty$. Then there exists g with $0 \leq g(x) \leq \rho(x)$ such that $\psi_g = |x|^{-1} * g$ satisfies $\psi_g(x) = V(x)$ a.e. when $\rho(x) - g(x) > 0$ and $\psi_g(x) \leq V(x)$ a.e.

Proof. Baxter proves this when ρ and g are measures. We give a simpler proof for functions. Consider $\mathcal{E}(g) \equiv D(g, g) - \int Vg$ and $E = \inf \{ \mathcal{E}(g) \mid 0 \leq g(x) \leq \rho(x) \}$. Let g^n be a minimizing sequence. There exists a subsequence that converges weakly in L^p to some g and, by Mazur's theorem (1933), there exists a sequence h^n of convex combinations of the g^n that converges strongly to g in L^p . Then a subsequence of the h^n converges a.e. to g . Clearly, $0 \leq h^n(x) \leq \rho(x)$. Since $\mathcal{E}(\cdot)$ is convex (this is crucial) $\mathcal{E}(h^n) \rightarrow E$ but, by dominated convergence, $\mathcal{E}(h) = \mathcal{E}(g)$. So g minimizes and satisfies (a.e.): $\psi_g(x) = \psi_{g^n}(x)$ when $0 < g(x) < \rho(x)$; $\psi_g(x) \leq V(x)$ when $g(x) = \rho(x)$ and $\rho(x) > 0$; $\psi_g(x) \geq V(x)$ when $g(x) = 0$ and $\rho(x) > 0$. We have to eliminate the possibility $\psi_g(x) - V(x) \equiv f(x) > 0$ when $g(x) = 0$. We claim $\psi_g \in \mathcal{D}$ and hence f is continuous and goes to zero at ∞ . Since $g \leq \rho$, $\psi_g \leq \psi_\rho$ so $\psi_g \rightarrow 0$ at infinity. To examine the continuity at $x=0$, write $\psi_g = h + (\psi_g - h)$ with $h = |x|^{-1} * (\chi g)$ and χ is the characteristic function of the ball $|x| < 1$. Clearly $\psi_g - h$ is continuous at $x=0$. Moreover, $\chi g \in L^p \cap L^1$ so $h \in \mathcal{D}$ by Lemma 3.1. (It is

here that $p > \frac{3}{2}$ is used.) Now, since $f \in \mathcal{D}$, $B = \{x \mid f(x) > 0\}$ is open and, since $x \in B \Rightarrow g(x) = 0$, f is subharmonic on B . But f vanishes on the boundary of B and at infinity, so B is empty. ■

IV. DEPENDENCE OF THE THOMAS-FERMI ENERGY ON THE NUCLEAR COORDINATES

In the previous sections TF theory was analyzed when the nuclear coordinates $\{R_i\}$ are held fixed. The one exception was Teller's theorem (Theorem 3.23) which states that the TF energy is greater than the TF energy for isolated atoms (which is the same as the energy when the R_i are infinitely far apart). Here, more detailed information about the dependence of E on the R_i is reviewed.

Note that in this section (and henceforth) E refers to the total energy, [Eq. (2.11)], including the repulsion U . This is crucial.

Although several unsolved problems remain, a fairly complete picture will emerge. The principal open problem is to prove the positivity of the pressure (Sec. IV.B) for subneutral molecules, and to prove it for deformations more general than uniform dilation. The results of this section have been proved only for TF theory, and it is not known which ones extend to the variants (see the discussion of TFD theory in Sec. IV.C).

A. The many-body potentials

The results here are from Benguria and Lieb, 1978a. As usual, the two-body atomic energy is *defined* to be the difference between the energy of a diatomic molecule (with nuclear separation R) and the energy of iso-

lated atoms. Teller's theorem states that this is always positive. We shall now investigate the k -body energy which can be defined similarly. The three-body energy will be shown to be negative, the four-body positive, etc. In all cases, only neutral systems will be considered; in this case there is a unique way to apportion the electron charge among the isolated atoms, namely, make them all neutral. An interesting problem is to treat the k -body energy for subneutral systems.

Definitions. When $c = \{c_1, c_2, \dots, c_k\}$ is a finite subset of the positive integers with $|c| = k$ elements, $E(c)$ denotes the TF energy for a neutral molecule consisting of nuclear charges $z_{c_i} > 0$ located R_{c_i} . $\phi(c, x)$ denotes the TF potential for this molecule. The z 's can all be different.

$$\varepsilon(c) = \sum_{b \subseteq c} (-1)^{|b|+|c|} E(b) \quad (4.1)$$

is the $|c|$ body energy for this molecule. Thus, if $c = \{1, 2\}$, $|c| = 2$ and the two-body energy is $\varepsilon(1, 2) = E(1, 2) - E(1) - E(2)$ as explained above. If $c = \{1, 2, 3\}$, $|c| = 3$ and the three-body energy is

$$\begin{aligned} \varepsilon(1, 2, 3) = & E(1, 2, 3) - [E(1, 2) + E(1, 3) + E(2, 3)] \\ & + E(1) + E(2) + E(3). \end{aligned}$$

$E(1), E(2), E(3)$ are atomic energies, of course. From Eq. (4.1)

$$E(c) = \sum_{b \subseteq c} \varepsilon(b). \quad (4.2)$$

It is worth remarking that the many-body energies (4.1) are defined in terms of the total energy E . It is equally possible to use $e = E - U$ on the right side of Eq. (4.1). e is the electronic contribution to E , so the corresponding ε 's would be the electronic contribution to the many-body potential. However, note that U contains only two-body pieces, $z_i z_j |R_i - R_j|^{-1}$. Therefore the two sets of ε 's agree whenever $|c| \geq 3$, i.e., the three- and higher-body ε 's are entirely electronic. As far as the two-body energy is concerned, $\varepsilon(1, 2)_{\text{tot}} > 0$ (Teller) but

$$\varepsilon(1, 2)_{\text{elec}} = \varepsilon(1, 2)_{\text{tot}} - U(1, 2) < 0$$

(Theorem 3.21).

In the following $b \subset c$ means b is a subset of c and $b \neq c$.

Theorem 4.1 (Sign of the many-body potential). *If c is not empty*

$$(-1)^{|c|} \varepsilon(c) > 0.$$

More generally, if $b \subset c$ and either $|c \setminus b| \geq 2$ or else $|b| = 0$ and $|c| > 0$

$$\tilde{E}(b, c) \equiv \sum_{b \subseteq a \subseteq c} (-1)^{|b|+|a|} E(a) > 0.$$

Theorem 4.2 (Remainder Theorem). *If $2 \leq \beta \leq |c|$ then the sign of*

$$E(c) - \sum_{\substack{b \subset c \\ |b| < \beta}} \varepsilon(b)$$

is $(-1)^\beta$. In other words, if, in Eq. (4.2), we sum only

over the terms smaller than β -body, the sign of the error is the sign of the first omitted terms.

Theorem 4.3 (Monotonicity of the many-body potential). *Suppose that $b \subset c$ and $|b| \geq 2$. Then*

$$(-1)^{|b|} \varepsilon(b) > (-1)^{|c|} \varepsilon(c).$$

Theorems 4.1 and 4.3 imply, for example,

$$0 > \varepsilon(1, 2, 3) > -\min[\varepsilon(1, 2), \varepsilon(1, 3), \varepsilon(2, 3)].$$

Theorem 4.4. *If $b \subset c$ and c is not empty*

$$\tilde{\phi}(b, c, x) \equiv \sum_{b \subseteq a \subseteq c} (-1)^{|a|+|b|} \phi(a, x) < 0.$$

Partial Proof. Basically Theorems 4.1, 4.2, and 4.3 are corollaries of Theorem 4.4 through the relation, for $j \in c$,

$$\frac{\partial E(c)}{\partial z_j} = \lim_{x \rightarrow R_j} \{ \phi(c, x) - z_j |x - R_j|^{-1} \}, \quad (\text{Theorem 2.13}).$$

As an illustration we shall prove here that $\varepsilon(1, 2, 3) < 0$; surprisingly, the proof is much more complicated when $|c| > 3$. The proof for $|c| = 3$ only uses that the function $(j^*)^{-1}$ is convex [cf. Eq. (3.1)]. The proof for $|c| > 3$ requires that $j(\rho) = \rho^k$ with $\frac{3}{2} \leq k \leq 2$.

First note that $\varepsilon(1, 2, 3) = 0$ when $z_3 = 0$. Thus it suffices to prove that $\partial \varepsilon(1, 2, 3) / \partial z_3 = F(R_3) < 0$, where

$$F(x) \equiv \phi(1, 2, 3, x) - \phi(1, 3, x) - \phi(2, 3, x) + \phi(3, x).$$

Now

$$\Delta F = 4\pi [\rho(1, 2, 3, x) - \rho(1, 3, x) - \rho(2, 3, x) + \rho(3, x)]$$

and $\rho = (\phi/\gamma)^{3/2}$. Let $B = \{x | F(x) > 0\}$. F is continuous, so B is open. We claim F is subharmonic on B , which implies B is empty. What is needed is the fact that $a - b - c + d \geq 0 \Rightarrow a^{3/2} - b^{3/2} - c^{3/2} + d^{3/2} \geq 0$ under the conditions that $a \geq b \geq d \geq 0$ and $a \geq c \geq d \geq 0$ (Theorem 3.4). But this is an elementary exercise in convex analysis. Finally, as in the strong form of Theorem 3.4, one can prove that F is strictly negative. ■

It is noteworthy that all the many-body potentials fall off at the same rate, R^{-7} . This will be shown in Sec. IV.C.

B. The positivity of the pressure

Teller's theorem (Theorem 3.23) suggests that the nuclear repulsion dominates the electronic attraction and therefore a molecule in TF theory should be unstable under local as well as global dilations.

Let us fix the nuclear charges $\underline{z} = \{z_1, \dots, z_k\}$ and move the R_i keeping λ fixed. Under which deformations does E decrease? We can also ask when $e = E - U$, the electronic contribution to the energy, decreases. A natural conjecture is the following: Suppose $R_i \rightarrow R'_i$ with $|R'_i - R'_j| \geq |R_i - R_j|$ for every pair i, j . Then

(i) E decreases and e increases.

(ii) Furthermore, if $\lambda_1 < \lambda_2$ then the decrease (increase) in $E(e)$ is smaller (larger) for λ_2 than for λ_1 . There is one case in which this conjecture can be proved; it is given in Theorem 4.7 due to Benguria (1981).

One interesting case is that of uniform dilation in

which each $R_i \rightarrow lR_i$. For this case we define the *pressure and reciprocal compressibility* to be

$$P(l) = - (3l^2)^{-1} dE(l)/dl \quad (4.3)$$

$$\kappa^{-1} = - (l/3) dP(l)/dl, \quad (4.4)$$

where $E(l)$ is the energy. This definition comes from thinking of the "volume" as proportional to l^3 . If $K(l)$ is the kinetic energy [Eq. (2.20)] then

$$3l^3 P(l) = E(l) + K(l).$$

To see this, define $E(\gamma, l)$ to be the energy with the parameter γ thought of as a variable (but with l fixed). Then, by setting $\rho(x, l) = l^{-3} \bar{\rho}(x/l, l)$, one easily sees that $E(\gamma, l) = l^{-1} E(\gamma/l, 1)$ and $K(\gamma, l) = l^{-1} K(\gamma/l, 1)$. Equation (4.4) follows from this and Theorem 2.13.

Note that Eq. (4.4) is true (for the same reason) in Q theory and also in TFD, TFW, and TFDW theories provided K is interpreted as Eq. (2.20) in TFD and as $(2.20) + \delta \int |\nabla \rho|^{1/2}{}^2$ in TFDW and TFW.

That $e = E - U$ increases under dilation has also been conjectured to hold in Q theory when $\lambda \leq Z$. It is known to hold for one electron, but an arbitrary number of nuclei (Lieb and Simon, 1978). There is one simple statement that can be made (in all theories): The (unique) minimum of e occurs when $l = 0$ (for any $\lambda > 0$), i.e., all the nuclei are at one point. To prove this, assume $R_1, \dots, R_k$ are not all identical and let ρ be the minimizing solution. Let $\psi = |x|^{-1} * \rho$. ψ has a maximum at some point R_0 . Now place all the nuclei at R_0 and use the same ρ as a variational ρ for this problem. Then, trivially, $e(R_0, \dots, R_0) < e(R_1, \dots, R_k)$, with the strict inequality being implied by the fact that this ρ does not satisfy the variational equation for $R_0, \dots, R_0$.

It is useful to have a formula for the variation of e with R_i . A natural extension of Theorem 2.13 (a "Feynman-Hellman"-type theorem) would be the following: Suppose $V_1, \dots, V_k \in \mathfrak{D}$ with

$$V_i(x) = \int dm_i(y) |y - x|^{-1} \quad (4.5)$$

and with m_i a positive measure of mass z_i . Take

$$V(x) = \sum_{i=1}^k V_i(x - R_i).$$

Then e is a C^1 function of the R_i and

$$\nabla_{R_i} e = \int \nabla V_i(x - R_i) \rho(x) dx = - \int dm_i(y) \nabla \psi(y + R_i), \quad (4.6)$$

with $\psi = |x|^{-1} * \rho$. Equation (4.6) is clearly true, and easy to prove if the m_i are suitably bounded. Benguria (unpublished) proved (4.6) when $V_i(x) = z_i |x|^{-1}$ for $|x| \geq a$ and $V_i(x) = z_i a^{-1}$ for $|x| \leq a$, with $a > 0$, i.e., $dm_i(y) = z_i(\text{const}) \delta(|y| - a)$. In this case, the last equality in Eq. (4.6) follows from LS, Lemma IV.4.

For point nuclei, on the other hand, (4.6) has not been proved; indeed, the quantities in (4.6) are not even well defined. We conjecture that the following is true when $V_i(x) = z_i |x|^{-1}$: e is a C^1 function of the R_i on the set where $R_i \neq R_j$, for all $i \neq j$, and

$$\nabla_{R_i} e = - z_i \lim_{a \downarrow 0} \int_{|x - R_i| > a} (x - R_i) |x - R_i|^{-3} \rho(x) dx \quad (4.7a)$$

$$= - \lim_{a \downarrow 0} \nabla_{R_i} \{ \psi(x) + (z_i/\gamma)^{3/2} (16\pi/3) |x - R_i|^{1/2} \}. \quad (4.7b)$$

Equation (4.7a) makes sense because, by Theorem 2.8,

$$\rho(x) = (z_i/\gamma)^{3/2} |x - R_i|^{-3/2} + O(|x - R_i|^{-1/2})$$

near R_i ; the angular integration over the first term vanishes. This leading term in ρ implies that near R_i , $\psi(x) \approx (\text{const}) - (z_i/\gamma)^{3/2} (16\pi/3) |x - R_i|^{1/2}$. The nondifferentiable, but spherically symmetric term in ψ is subtracted in Eq. (4.7b).

The following theorems have been proved so far. (Theorems 4.5 and 4.6 are in Benguria and Lieb, 1978b; Theorem 4.7 is in Benguria, 1981.)

Theorem 4.5 (Uniform dilation). *Replace each R_i by lR_i and call the energy $E(\lambda, l)$. If $\lambda = Z$ then $E(\lambda, l)$ is strictly monotone decreasing and convex in l . In particular, the pressure and compressibility are positive.*

Remarks. (i) If $\lambda = 0$ the conclusion is obviously also true. In Benguria and Lieb (1978b) it is conjectured that this theorem holds for all λ . That $e = E - U$ is monotone increasing is also conjectured there.

(ii) In Benguria and Lieb (1978b) several interesting subadditivity and convexity properties of the energy and potential are also proved.

Theorem 4.6 (Molecule with planar symmetry). *Suppose the molecule is symmetric with respect to the plane $P = \{(x^1, x^2, x^3) | x^1 = 0\}$ and suppose no nucleus lies in the plane. Neutrality is not assumed. Let R_i^1 denote the 1 coordinate of nucleus i and, for all i , replace R_i^1 by $R_i^1 \pm l$, with $\pm$ if $R_i^1 \geq 0$, and $l \geq 0$. Then for all fixed $\lambda \leq Z$, E is decreasing in l .*

Remark. For a homopolar diatomic molecule the dilations in Theorems 4.5 and 4.6 are the same. Balazs (1967) first proved Theorem 4.6 in this case. For a general diatomic molecule, Benguria's Theorem 4.7 is the strongest theorem.

Theorem 4.7. *Suppose there exists a plane P containing $R_1, \dots, R_m$ and such that all the other R_j (with $j = m+1, \dots, k$) are on one (open) side of P (call this side P^*). Assume the nuclei at $R_1, \dots, R_m$ are point nuclei, but the nuclei at $R_{m+1}, \dots, R_k$ are anything in $\mathfrak{D}$ and given by Eq. (4.5) with the supports of $m_i \in P^*$ (this includes point nuclei). Let n be the normal to P pointing away from P^* . Let $l_1, \dots, l_m \geq 0$ be given and let $R_i \rightarrow R_i + l_i n$ for $i = 1, \dots, m$. Let $E(\lambda, l)$ denote the energy for fixed $\lambda \leq Z$ and let $\Delta E(\lambda, l) = E(\lambda, l) - E(\lambda, 0)$ denote the change in energy. Likewise define $\Delta e(\lambda, l) = \Delta E(\lambda, l) - \Delta U$. Then*

- (i) $\Delta e(\lambda, l) \geq 0$,
- (ii) $\Delta E(\lambda, l) \leq 0$,
- (iii) $\Delta E(\lambda_1, l) \leq \Delta E(\lambda_2, l)$ if $\lambda_1 \leq \lambda_2$,
- (iv) $\Delta e(\lambda_1, l) \leq \Delta e(\lambda_2, l)$ if $\lambda_1 \leq \lambda_2$.

To prove Theorem 4.7 the following Lemma 4.8, which is of independent interest, is needed.

Lemma 4.8. *Assume the plane P , with $R_1, \dots, R_m$ in P and $R_{m+1}, \dots, R_k$ in P^* as in Theorem 4.7. However, point nuclei are not assumed. Instead, assume each V_i*

$\in D$ and given by Eq. (4.5), with m_i required to be spherically symmetric for $i=1, \dots, m$. This includes point nuclei. Assume also that the support of $m_i \subset P^*$ for $i=1, \dots, m$. If $x \in P^*$ then x^* is defined to be the reflection of x through P . Let ϕ be the potential. For $x \in P^*$, let $\phi_+(x) = \phi(x^*)$ and $f(x) = \phi(x) - \phi_-(x)$. Then

- (i) $f(x) > 0$ for $x \in P^*$.
- (ii) For each $x \in P^*$, $f(x)$ strictly decreases when λ increases.
- (iii) $\rho(x) - \rho(x^*) \geq 0$ for $x \in P^*$.

Question. Is it true that $\rho(x) - \rho(x^*)$ is a monotone increasing function of λ ?

Proof. (i) Clearly $f(x) = 0$ on $\partial P^* = P$ and at ∞ . Let $B = \{x \in P^* | f(x) < 0\}$. Since each $V_i(x)$ is symmetric decreasing the singularities of V are not in B . Thus B is open. On B , $-\Delta f(x)/4\pi \geq -\rho(x) + \rho(x^*) > 0$. Thus f is superharmonic on B so B is empty. By the strong maximum principle $f(x) > 0$, in fact, for $x \in P^*$.

(ii) Let $\lambda' < \lambda$ with corresponding f' and f . We want to prove $B = \{x \in P^* | f(x) - f'(x) > 0\}$ is empty. B is open and $f - f' = 0$ on P and at ∞ . $\Delta(f - f')/4\pi = a^{3/2} - b^{3/2} - c^{3/2} + d^{3/2} \equiv h$, where $a = \phi - \mu$, $b = \phi' - \mu'$, $c = \phi_- - \mu$, $d = \phi'_- - \mu'$. By (i) and Corollary 3.8, $a \geq b > d$ and $a > c \geq d$ for all $x \in P^*$. In B , $a + d > b + c$. Thus $h \geq 0$ in B , whence $f - f'$ is subharmonic on B and hence B is empty. Again, one can prove the stronger result that $f - f' < 0$ for $x \in P^*$. Trivially, (i) $\Rightarrow$ (iii) through the TF equation. ■

Proof of Theorem 4.7. We may assume all the l_i are equal to some common l , for otherwise if $l_1 \leq l_2 \leq \dots \leq l_m$ we could first move all the m nuclei by l_1 , then move $R_2, \dots, R_m$ by $l_2 - l_1$, etc. Next, replace all the point nuclei at $R_1, \dots, R_m$ by smeared potentials given by Eq. (4.5) with $dm_i(x) = z_i g^{(m)}(x) dx$ where $g^{(n)}(x) \in C_0^\infty$ and $g^{(n)}$ is symmetric decreasing and with sufficiently small support such that the supports of dm_i ($i=1, \dots, m$) are pairwise disjoint and also disjoint from the supports of dm_i ($i=m+1, \dots, k$). Under these conditions, e is C^1 in $R_1, \dots, R_m$ in some neighborhood of the original $R_1, \dots, R_m$ with derivatives given by Eq. (4.6). We shall prove

- (i)' $\mathbf{n} \cdot \nabla_{R_i} e \geq 0$,
- (ii)' $\mathbf{n} \cdot \nabla_{R_i} E \leq 0$,

and that (iii) and (iv) hold for these derivatives. Then the theorem is proved because the original point potentials $z_i |x|^{-1}$ can be approximated in $L^{5/2}$ norm by these smeared potentials $z_i |x|^{-1} * g^{(n)}$, and the energies $e^{(n)}$ and $E^{(n)}$ converge to e and E by LS, Theorem II.15. If (i)' holds for $e^{(n)}$, then $(d/dl)e^{(n)}(\lambda, l) \geq 0$ with $R_i - R_i + l\mathbf{n}$, $i=1, \dots, m$, and, by integration, (i) holds for $e^{(n)}$. Then, when $n \rightarrow \infty$, (i) holds for e . The same applies to (ii)-(iv). Henceforth the superscript (n) will be suppressed.

Assume $\mathbf{n} = (1, 0, 0)$, $P = \{x | x^1 = 0\}$, and thus $(R_i)^1 = 0$ for $i=1, \dots, m$. Since g is symmetric decreasing,

$$(\partial g / \partial x^1)(x^1, x^2, x^3) = -x^1 h(x^1, x^2, x^3)$$

with $h(x) \geq 0$ and

$$h(x^1, x^2, x^3) = h(-x^1, x^2, x^3).$$

Likewise,

$$(\partial V_i / \partial x^1)(x^1, x^2, x^3) = -z_i x^1 p(x^1, x^2, x^3),$$

and p has the same properties as h . To prove (i)' use Eq. (4.6) whence

$$\begin{aligned} \mathbf{n} \cdot \nabla_{R_i} e / z_i &= - \int x^1 p(x - R_i) \rho(x) dx \\ &= - \int_{x^1 \leq 0} p(x - R_i) [\rho(x) - \rho(x^*)] x^1 dx \geq 0 \end{aligned}$$

by Lemma 4.8. To prove (ii)' use the second integral in Eq. (4.6), whence

$$B_i \equiv \mathbf{n} \cdot \nabla_{e_i} E = \int dm_i(y) \mathbf{n} \cdot \nabla \phi(y + R_i),$$

where ϕ is the potential. [Note: $V_i(x - R_i)$ is symmetric in x about R_i , so the term $\nabla V_i(x - R_i)$ does not contribute to this integral.] Since V_i is C^∞ it is easy to see that ϕ is also C^∞ near R_i . Now integrate by parts:

$$\begin{aligned} B_i &= - \int \mathbf{n} \cdot \nabla g(y) \phi(y + R_i) dy \\ &= \int y^1 h(y) \phi(y + R_i) dy \\ &= \int_{y^1 \leq 0} y^1 h(y) [\phi(y + R_i) - \phi_-(y + R_i)] dy \leq 0 \end{aligned}$$

by Lemma 4.8. To prove (iii) note that the last quantity [] decreases when λ increases by Lemma 4.8. Clearly (iii) is equivalent to (iv). ■

Proof of Theorem 4.6. Let $\rho(x)$ be the density when $l=0$. For $l > 0$ use the variational $\tilde{\rho}$ given by $\tilde{\rho}(x^1, x^2, x^3) = \rho(x^1 \mp l, x^2, x^3)$ if $x^1 \geq l$ and $\tilde{\rho}(x) = 0$ otherwise. Then all terms in the energy $\mathcal{G}(\tilde{\rho})$ remain the same except for the Coulomb interaction of the two charge distributions on either side of the plane P . This term is of the form

$$W(l) = \int_{x^1, y^1 \geq 0} d^3x d^3y f(x) f(y) \times [(x^1 + y^1 + 2l)^2 + (x^2 - y^2)^2 + (x^3 - y^3)^2]^{-1/2},$$

where $f(x) = -\rho(x) + \sum' z_i \delta(x - R_i)$ and the $\sum'$ is over those R_i with $R_i^1 > 0$. Since the Coulomb potential is reflection positive (Benguria and Lieb, 1978, Lemma B.2), $W(l)$ is a decreasing, log convex function of l . ■

Proof of Theorem 4.5. Let $\underline{z} = (z_1, \dots, z_k)$ and write $E(\underline{z})$, $K(\underline{z})$, $A(\underline{z})$, and $R(\underline{z})$ for the energy and its components (cf. Sec. II.E) of a neutral molecule. These functions are defined on $\mathbb{R}_+^k$. For an atom $3P = E + K = 0$ (Theorem 2.14). By Theorem 3.23, $E \geq \sum_1^k E^{\text{atom}}(z_i)$ and, by Theorem 4.10, $K \geq \sum_1^k K^{\text{atom}}(z_i)$. This shows $P \geq 0$. Likewise, by Theorem 4.12, $\kappa^{-1} > 0$ and $E(\underline{z}, l)$ is convex in l (equivalently $l^2 P$ is decreasing in l). ■

Definition. Let f be a real valued function on $\mathbb{R}_+^k$ and $\underline{z}_1, \underline{z}_2, \underline{z}_3 \in \mathbb{R}_+^k$. Then f is

- (i) weakly superadditive (WSA) $\Leftrightarrow f(\underline{z}_1 + \underline{z}_2) \geq f(\underline{z}_1) + f(\underline{z}_2)$ whenever $(z_1)_i (z_2)_i = 0$, all i ,
- (ii) superadditive (SA) $\Leftrightarrow f(\underline{z}_1 + \underline{z}_2) \geq f(\underline{z}_1) + f(\underline{z}_2)$,
- (iii) strongly superadditive (SSA) $\Leftrightarrow f(\underline{z}_1 + \underline{z}_2 + \underline{z}_3) + f(\underline{z}_1) \geq f(\underline{z}_1 + \underline{z}_2) + f(\underline{z}_1 + \underline{z}_3)$.

Theorems 4.9-4.12 are for neutral molecules.

Theorem 4.9. As a function of $\underline{z} \in \mathbb{R}_+^k$, for each fixed

$x \in \mathbb{R}^3$,

(i) $-\phi(\underline{z}, x)$ is SSA, convex, and decreasing (the latter is Teller's lemma),

(ii) $\phi(\underline{z}, x) \in C^1(\mathbb{R}^3)$ and $\in C^2(\mathbb{R}^3 \setminus 0)$,

(iii) $\phi_i(\underline{z}, x)$ is decreasing in $\underline{z}$ and > 0 ,

(A subscript i denotes $\partial/\partial z_i$.)

(iv) $\phi_{ij}(\underline{z}, x) \leq 0$ (all i, j) and is negative semidefinite as a $k \times k$ matrix.

Remark. It is easy to prove that when $f \in C^2(\mathbb{R}^3)$ then SSA is equivalent to $f_{ij} \geq 0$ for all i, j . See Benguria and Lieb, 1978b, for this and similar equivalences.

Theorem 4.10. $K(\underline{z}) \in C^1(\mathbb{R}^3)$ and $\in C^2(\mathbb{R}^3 \setminus 0)$

$$(i) K_i(\underline{z}) = 3 \lim_{\underline{z} \rightarrow \underline{R}_i} \left\{ \phi(\underline{z}, x) - \sum_{j=1}^k z_j \phi_j(\underline{z}, x) \right\},$$

$$(ii) K_{ij}(\underline{z}) = -3 \sum_{p=1}^k z_p \phi_{ij}(\underline{z}, R_p),$$

(iii) $K(\underline{z})$, $R(\underline{z})$, and $A(\underline{z})$ are SSA and SA and convex,

(iv) $E(\underline{z})$ is WSA (Teller's theorem).

Definition. $X(\underline{z}) \equiv 3K(\underline{z}) - \sum_{i=1}^k z_i K_i(\underline{z})$.

Theorem 4.11. $X(\underline{z})$ is SA and SSA and ray convex.

I.e., $X(\lambda \underline{z}_1 + (1-\lambda)\underline{z}_2) \leq \lambda X(\underline{z}_1) + (1-\lambda)X(\underline{z}_2)$, $0 \leq \lambda \leq 1$, when $\underline{z}_1, \underline{z}_2 \in \mathbb{R}^3$ and either $\underline{z}_1 - \underline{z}_2$ or $\underline{z}_2 - \underline{z}_1 \in \mathbb{R}^3$.

Theorem 4.12. (i) $3l^3 P = E + K$,

(ii) $9l^3 \kappa^{-1} = 6l^3 P + 2E + 3X$,

(iii) P and κ^{-1} are WSA and non-negative,

(iv) $l^2 P$ is decreasing in l . Equivalently, E is convex in l . Equivalently, $2E + 3X \geq 0$

[note: $\partial(l^2 P)/\partial l = 2lP - 3l\kappa^{-1} = -(\frac{1}{3})(2E + 3X)$].

Proof of (iv). $2E + 3X = 0$ for an atom. By Theorem 4.10, $2E + 3X \geq 0$. ■

The proofs of Theorems 4.9–4.12 are complicated. However, if all necessary derivatives are assumed to exist, then an easy heuristic proof can be given (see Benguria and Lieb, 1978b). We illustrate this for K being SSA, which is equivalent to $K_{ij} \geq 0$, all i, j . This will then prove $P \geq 0$, since $K(0) = 0$. First we show $\phi_{ij} \leq 0$ and then Theorem 4.10 (ii).

Differentiate the TF differential equation $[\Delta\phi/4\pi = -\sum z_j \delta(x - R_j) + (\phi/\gamma)^{3/2}]$, which holds for any neutral system] with respect to z_i and then z_j :

$$\mathcal{L}\phi_i = \delta(x - R_i), \quad (4.8)$$

$$\mathcal{L}\phi_{ij} = -(3/4\gamma^{3/2})\phi^{-1/2}\phi_i\phi_j, \quad (4.9)$$

with $\mathcal{L} = -\Delta/4\pi + (3\gamma^{-3/2}/2)\phi(x)^{1/2}$. The kernel for $\mathcal{L}^{-1}$ is a positive function, so $\phi_i \geq 0$. Likewise $\phi_{ij} \leq 0$ and ϕ_{ij} is a negative semidefinite matrix.

Next, $K = (3\gamma^{-3/2}/5)\int \phi^{3/2}$, so

$$K_{ij} = (3\gamma^{3/2}/2) \left\{ \int \phi^{3/2}\phi_{ij} + \frac{3}{2} \int \phi^{1/2}\phi_i\phi_j \right\}.$$

Using Eq. (4.9) and integrating by parts,

$$\begin{aligned} K_{ij} &= 3 \int \phi_{ij} [\Delta\phi/4\pi - (\phi/\gamma)^{3/2}] \\ &= -3 \sum_{p=1}^k z_p \phi_{ij}(R_p) \geq 0. \end{aligned}$$

C. The long-range interaction of atoms

In Sec. IV.B it was shown that the energy of a molecule decreases monotonically under dilation (at least for neutral molecules). If the $R_i \rightarrow lR_i$ then, for small l , E is dominated by U , so $E \approx l^{-1}$. To complete the picture it is necessary to know what happens for large l . We define

$$\Delta E = E^{\text{mol}} - \sum_{i=1}^k E^{\text{atom}}. \quad (4.10)$$

For large l it is reasonable to consider only neutral molecules, for otherwise $\Delta E \approx l^{-1}$ because of the unscreened Coulomb interaction. In the neutral case $\Delta E \approx l^{-7}$, as proved by Brezis and Lieb (1979). This result (l^{-7}) is not easy to ascertain numerically (Lee, Longmire, and Rosenbluth, 1974), so once again the importance of pure analysis in the field is demonstrated. Some heuristic remarks about the result are given at the end of this section.

A surprising result is that all the many-body potentials are $\approx l^{-7}$. Thus in TF theory it is not true that the interaction of atoms may be approximated purely by pair potentials at large distances.

An interesting open problem is to find the long-range interaction of polyatomic molecules of fixed shape. Presumably this is also $\approx l^{-7}$.

Theorem 4.13. For a neutral molecule, let the nuclear coordinates be lR_i with $\{R_i, z_i\} = (\underline{R}, \underline{z})$ fixed and $z_i > 0$. Then

$$\Delta E(l, \underline{z}, \underline{R}) \approx l^{-7} C(l, \underline{z}, \underline{R}),$$

where C is increasing in l and has a finite limit, $\Gamma(\underline{R}) > 0$ as $l \rightarrow \infty$. Γ is independent of $\underline{z}$. Furthermore, if A denotes a subset of the nuclei (with coordinates $\underline{R}_A$), and $\varepsilon(A)$ is the many-body potential of Eq. (4.1), then, by (4.1), for $|A| \geq 2$

$$l^7 \varepsilon(A) \rightarrow \sum_{B \subseteq A} (-1)^{|A|-|B|} \Gamma(\underline{R}_B) \quad (4.11)$$

and the right side of Eq. (4.11) is strictly positive (negative) if $|A|$ is even (odd).

Proof of first part. By scaling, Eq. (2.24), we find that

$$\Delta E(l, \underline{z}, \underline{R}) = l^{-7} \{ E(l^3 \underline{z}, \underline{R}) - \sum_j E^{\text{atom}}(l^3 z_j) \}.$$

Therefore, C increasing is equivalent to $f = E^{\text{mol}} - \sum E^{\text{atom}}$ increasing in $\underline{z}$. But $\partial f / \partial z_j = \lim_{\underline{x} \rightarrow R_j} \phi^{\text{mol}}(\underline{x}) - \phi^{\text{atom}}(\underline{x})$, and this is positive by Teller's lemma. All that has to be checked is that C is bounded above. This is done by means of a variational ρ for E^{mol} . Let B_i be a ball of radius $l\gamma_i$ centered at lR_i ; the γ_i are chosen so that the B_i are disjoint. Let $\rho_i(\underline{x}) = \rho^{\text{atom}}(\underline{x} - lR_i)$ be the TF atomic densities for z_i , and let $\rho(\underline{x}) = \rho_i(\underline{x})$ in B_i and $\rho(\underline{x}) = 0$ otherwise. Of course $\int \rho < \sum z_j$ but this is immaterial for a variational calculation, since the minimum molecular energy occurs when $\int \rho = \sum z_j$. It is easy to check that $f < (\text{const})l^{-7}$. Finally, since f is monotone in each z_i , $\lim_{l \rightarrow \infty} f$ must be independent of the z_j . ■

Remarks. (i) The variational calculation shows clearly why Γ is independent of the z_j . The long-range interaction comes, in some sense, from the tails of the atomic ρ 's, but these tails are independent of z , namely $\rho(x) \approx (3\gamma/\pi)^3 |x|^{-6}$. (See Theorem 2.10.)

(ii) At first sight it might appear counterintuitive that the interaction is $+l^{-7}$ and not $-l^{-6}$, as would be obtained from a dipole-dipole interaction. The following heuristic remark might be useful in this respect. Consider two neutral atoms separated by a large distance R . In the quantum theory, as in all the theories discussed in this paper, there is almost no static polarization of the atoms; i.e., there is no polarization of the single-particle density ρ . TF theory is therefore correct as far as the density is concerned. The reason there is no polarization is that the formation of a dipole moment d increases the atomic energy by $+ad^2$ with $\alpha > 0$. The dipole-dipole energy gain is $-(\text{const})d^2R^{-3}$. Hence, if R is large enough, the formation of dipoles does not decrease the energy. In quantum theory there is, in fact, a $-R^{-6}$ dipolar energy, but this effect is a correlation, and not a static effect. There are two ways to view it. In second-order perturbation theory there are *virtual* transitions to excited, polarized states. Alternatively, the electrons in each atom are correlated so that they go around their respective atoms in phase, but spherically symmetrically. This correlated motion increases the internal atomic energy only by ad^4 , not d^2 . In short, the $-R^{-6}$ interaction arises from the fact that the density ρ is not that of a structureless "fluid" but is the average density of many separate particles which can be correlated. This fact poses a serious problem for any "density functional approach." It is necessary to predict a $-R^{-6}$ dipolar interaction, yet predict essentially zero static polarization.

An explicit formula for $\Gamma(R)$ does not seem to be easy to obtain. Two not very explicit formulas are given in Brezis and Lieb, 1979. One is simply to integrate the formula for $\partial f/\partial l = 3l^2 \sum_j \partial f/\partial z_j$ given in the above proof. Another is obtained by noting that Γ is related to ϕ in the limit $z \rightarrow \infty$. This limiting ϕ can be defined, and satisfies the TF differential equation, but with a *strong singularity* at R_j instead of the usual $z |x - R_j|^{-1}$ singularity. As we saw in Theorem 2.11, the only other singularity allowed for the TF equation is $\phi(x) \approx \gamma^3 (3/\pi)^2 |x - R_j|^{-4}$. Therefore that peculiar solution to the TF equation does have physical interest; it is related to the asymptotic behavior of the interatomic interaction.

TFD theory. Here the interaction for large l is *precisely zero* and not l^{-7} . To be precise, $\Delta E = 0$ when the spacing between each pair $|R_i - R_j|$ exceeds a critical length, $L(z_i) + L(z_j)$. The same is *a fortiori* true for the many-body potentials ε .

The reason is the following. In TFD theory an atomic ρ has compact support, namely a ball of radius $L(z)$. See Theorem 6.6. When $|R_i - R_j| > L(z_i) + L(z_j)$, then $\rho(x) = \sum_j \rho_j(x - R_j; z_j)$ where $\rho(\cdot; \cdot)$ is the TFD atomic ρ . Since each atom is neutral, there is then no residual interaction, by Newton's theorem. One may question whether the ρ just defined is correct. It is trivial to check that it satisfies the TFD equation and, since the solution is unique, this must be the correct ρ .

V. THOMAS-FERMI THEORY AS THE $Z \rightarrow \infty$ LIMIT OF QUANTUM THEORY

Our goal in this section is to show that TF theory is the $Z \rightarrow \infty$ limit of Q theory and that it correctly describes the cores of heavy atoms. This is the perspective from which to view TF theory, and in this light it is seen to be a cornerstone of many-body theory, just as the theory of the hydrogen atom is an opposite cornerstone useful for thinking about light atoms. We shall not review the stability of matter question here (see Lieb, 1976).

In units in which $\hbar^2/2m = 1$ and $|e| = 1$ the Hamiltonian for N electrons is

$$H_N = \sum_{i=1}^N \{-\Delta_i + V(x_i)\} + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} + U. \quad (5.1)$$

E_N , $\rho_N(x)$, and μ will denote the TF energy, ρ and μ corresponding to this problem with $\lambda = N$ electrons if

$$N \leq Z = \sum_{j=1}^k z_j.$$

Of course, γ is taken to be γ_p [see Eq. (2.6)]. If $N > Z$ then these quantities are *defined* to be the corresponding TF quantities for $N = Z$. E_N^Q denotes the ground-state energy of H_N (defined to be $\inf \text{spec} H_N$) on the physical Hilbert space $\mathcal{H}_N = \wedge^N L^2(\mathbb{R}^3; \mathbb{C})$ (antisymmetric tensor product). q is the number of spin states ($= 2$ for electrons), but it is convenient to have it arbitrary, but fixed. The TF quantities also depend on q through γ_p .

A. The $Z \rightarrow \infty$ limit for the energy and density

Let us first concentrate on the energy; later on we shall investigate the meaning of $\rho(x)$. For simplicity the number of nuclei is fixed to be k ; it is possible to derive theorems similar to the following if $k \rightarrow \infty$ in a suitable way (e.g., a solid with periodically arranged nuclei), but we shall not do so here. In TF theory the relevant scale length is $Z^{-1/3}$ and therefore we shall consider the following limit.

Fix $\{z^0, R^0\} = \{z_j^0, R_j^0\}_{j=1}^k$ and $\lambda > 0$. For each $N = 1, 2, \dots$, define a_N by $\lambda a_N = N$, and in H_N , replace z_j by $a_N z_j^0$ and R_j by $a_N^{-1/3} R_j^0$. Thus $\lambda = Z^0 N/Z$, and a_N is the scale parameter. The TF quantities scale as [Eq. (2.24)]:

$$\begin{aligned} E_{\lambda a}(\underline{a} z^0, a^{-1/3} \underline{R}^0) &= a^{7/3} E_{\lambda}(\underline{z}^0, \underline{R}^0), \\ \rho_{\lambda a}(a^{-1/3} x, \underline{a} z^0, a^{-1/3} \underline{R}^0) &= a^2 \rho_{\lambda}(x, \underline{z}^0, \underline{R}^0). \end{aligned} \quad (5.2)$$

In this limit the nuclear spacing decreases as $a_N^{-1/3} \sim N^{-1/3} \sim Z^{-1/3}$. This should be viewed as a refinement rather than as a necessity. If instead the R_j are fixed $= R_j^0$, then in the limit one has isolated atoms. All that really matters are the limits $N^{1/3} |R_i - R_j|$.

Theorem 5.1 (LS Sec. III). *With $N = \lambda a_N$ as above*

$$\lim_{N \rightarrow \infty} a_N^{7/3} E_N^Q(a_N \underline{z}^0, a_N^{-1/3} \underline{R}^0) = E_{\lambda}(\underline{z}^0, \underline{R}^0).$$

The proof is via upper and lower bounds for E_N^Q . The upper bound is greater than the Hartree-Fock energy, which therefore proves that Hartree-Fock theory is correct to the order we are considering, namely $N^{1/3}$.

1. Upper bound for E_N^Q

The original LS proof used a variational calculation with a determinantal wave function; this is cumbersome. Baumgartner (1976) gave a simpler proof (both upper and lower bounds) which intrinsically relied on the same Dirichlet-Neumann bracketing ideas as in LS. Here, we give a new upper bound (Lieb, 1981a) that uses coherent states; these will also be very useful for obtaining a lower bound.

Let $y = (x, \sigma)$ denote a single space-spin pair and $\int dy = \sum_{\sigma=1}^2 \int dx$. Let $K(y, y')$ be any *admissible single-particle density matrix* for N fermions, namely $0 \leq K \leq I$ [as an operator on $L^2(\mathbb{R}^3; \mathbb{C}^2)$] and $\text{Tr} K = N$. Let h be the single-particle operator $-\Delta + V(x)$. Then (Lieb, 1981a)

$$E_N^Q \leq E_N^{\text{HF}} \leq \tilde{E}(K), \quad (5.3)$$

with

$$\begin{aligned} \tilde{E}(K) = & \text{Tr} K h + \frac{1}{2} \iint dy dy' |x - x'|^{-1} \\ & \times \{K(y, y')K(y', y') - |K(y, y')|^2\}. \end{aligned} \quad (5.4)$$

In Eq. (5.3), E_N^{HF} is the Hartree-Fock energy. Since $|x - x'|^{-1}$ is positive we can drop the "exchange term," $-|K|^2$, in Eq. (5.4) for the purposes of an upper bound.

First, suppose $N \leq Z$. To construct K , let $g(x)$ be any function on $\mathbb{R}^3$ such that $\int |g|^2 = 1$ and let $M(p, r)$ be any function on $\mathbb{R}^3 \times \mathbb{R}^3$ such that $0 \leq M(p, r) \leq 1$ and $(2\pi)^{-3} \times \int M dp dr = N/q$. Then the coherent states in $L^2(\mathbb{R}^3)$ which we shall use are

$$f_{p,r}(x) = g(x - r) \exp[ip \cdot x] \quad (5.5)$$

and

$$\begin{aligned} K(y, y') = & I_q (2\pi)^{-3} \int dp dr g(x - r) g(x' - r) M(p, r) \\ & \times \exp[ip \cdot (x - x')]. \end{aligned} \quad (5.6)$$

I_q is the identity operator in spin space. It is easy to check that $\text{Tr} K = N$ and that for any normalized ϕ in L^2 , $(\phi, K\phi) \leq 1$ by using Parseval's theorem and the properties of g and M . Thus K is admissible.

We choose [with $\rho = \rho_{\min}(V, Z)$ in Eqs. (5.7)–(5.26)]

$$M(p, r) = \theta(\gamma_p \rho(r)^{2/3} - p^2), \quad (5.7)$$

where $\theta(t) = 1$ if $t \geq 0$ and $\theta(t) = 0$ otherwise. γ_p is given in Eq. (2.6). One easily computes

$$K(y, y) = q^{-1} I_q \rho_g(x), \quad (5.8)$$

$$\text{Tr}(-\Delta)K = (3\gamma_p/5) \int \rho(x)^{5/3} dx + N \int |\nabla g(x)|^2 dx, \quad (5.9)$$

$$\text{Tr} VK = \int V_g(x) \rho(x) dx, \quad (5.10)$$

where $\rho_g = |g|^2 * \rho$ and $V_g = V * |g|^2$.

For $g(x)$ we choose

$$g(x) = (2\pi R)^{-1/2} |x|^{-1} \sin(\pi|x|/R) \quad (5.11)$$

for $|x| \leq R$, and $g = 0$ otherwise, and with $R = N^{2/5} Z^{-1}$.

Then

$$\int |\nabla g|^2 = \pi^2/R^2 = \pi^2 Z^2 N^{-4/5}.$$

The electron-electron interaction term in Eq. (5.4) is less than $D(\rho, \rho)$ because, as an operator (and function),

$$[|g|^2 * |x|^{-1} * |g|^2](x - x') < |x - x'|^{-1}.$$

To see this, use Fourier transforms. Thus

$$\begin{aligned} E_N^Q \leq \tilde{E}(K) \leq E_N + \pi^2 N^{1/5} Z^2 \\ + \int [V(x) - V_g(x)] \rho(x) dx. \end{aligned} \quad (5.12)$$

To bound the last term in Eq. (5.12) note that, by Newton's theorem, $|x|^{-1} - |g|^2 * |x|^{-1} = 0$ for $x \geq R$. Furthermore, with the scaling we have employed, $|R_i - R_j| > 2R$ for all $i \neq j$ and N large enough. Since $\gamma_p \rho^{2/3}(x) < V(x)$, then for sufficiently large N and for $|x - R_i| \leq R$ we have $\gamma_p \rho(x)^{2/3} < 2z_i |x - R_i|^{-1}$. Thus the last integral in Eq. (5.12) is bounded above for large N by

$$3\gamma_p^{-3/2} \sum_{j=1}^k z_j^{5/2} A$$

with

$$A = \int_{|x| \leq R} |x|^{-5/2} dx = 8\pi R^{1/2} = 8\pi N^{1/5} Z^{-1/2}.$$

If $N \leq Z$, we have established an adequate upper bound, namely,

$$E_N^Q - E_N \leq (\text{const}) N^{1/5} Z^2. \quad (5.13)$$

Since $Z \approx N$, this error is $\approx N^{11/5}$, and this is small compared to E , which is $\approx N^{7/3}$.

If $N > Z$ we use $K = K^1 + K^*$ where K^1 is given above (with $N = Z$) and K^* is a density matrix (really, a sequence of density matrices) whose trace is $N - Z$ and whose support is a distance d arbitrarily far away from the origin. K^* does not contribute to $\tilde{E}(K)$ in the limit $d \rightarrow \infty$. ■

2. Lower bound for E_N^Q

In LS a lower bound was constructed by decomposing $\mathbb{R}^3$ into boxes and using Neumann boundary conditions on these boxes. However, control of the singularities of V caused unpleasant problems. Here we use coherent states again (cf. Thirring, 1981).

Let $\psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)$ be any normalized function in $\mathcal{H}_N$ and let

$$\rho_g(x) = N \sum_{\sigma=1}^q \int |\psi(x, x_2, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 dx_2 \cdots dx_N \quad (5.14)$$

$$E_g = (\psi, H_N \psi) \quad (5.15)$$

$$T_g = \left(\psi, - \sum \Delta_i \psi \right). \quad (5.16)$$

It is known that (Lieb, 1979; Lieb and Oxford, 1981)

$$\begin{aligned} I_g = & \left(\psi, \sum_{i,j} |x_i - x_j|^{-1} \psi \right) \\ \geq & D(\rho_g, \rho_g) - (1.68) \int \rho_g(x)^{4/3} dx. \end{aligned} \quad (5.17)$$

Choose any $\bar{\rho}(x) \geq 0$ and $\bar{\phi} = V - |x|^{-1} * \bar{\rho}$. Since $D(\rho_g - \bar{\rho}, \rho_g - \bar{\rho}) \geq 0$, we have for any $0 \leq \varepsilon < 1$

$$E_\theta \geq \left(\psi, \sum_{i=1}^N h_i \psi \right) + U - D(\tilde{\rho}, \tilde{\rho}) - (1.68) \int \rho_\theta^{4/3} + \varepsilon T_\theta, \quad (5.18)$$

with

$$h = -(1 - \varepsilon)\Delta - \tilde{\phi}(x) \quad (5.19)$$

being a single-particle operator.

We shall consider $\tilde{\rho}$ to be the TF density for the problem with γ_ρ replaced by $(1 - \varepsilon)\gamma_\rho$, and with the same $\lambda = \min(N, Z)$. $-\tilde{\mu}$ and $\tilde{E}$ are the corresponding chemical potential and energy.

Let $f_{p,r}$ be the coherent states in $L^2(\mathbb{R}^3)$ given by Eq. (5.5) and $\pi_{p,r}$ (projection onto $f_{p,r}$) $\otimes I_G$. For any function $m(y) = m(x, \sigma)$ in $L^2(\mathbb{R}^3; \mathbb{C}^N)$ we easily compute:

$$\begin{aligned} (m, m) &= (2\pi)^{-3} \int dp dr (m, \pi_{p,r} m), \\ \int |\nabla m|^2 dz &= (2\pi)^{-3} \int dp dr p^2 (m, \pi_{p,r} m) \\ &\quad - (m, m) \int |\nabla g(x)|^2 dx, \\ \int |m|^2 \tilde{\phi}_\theta(x) dz &= (2\pi)^{-3} \int dp dr \tilde{\phi}(r) (m, \pi_{p,r} m), \quad (5.20) \end{aligned}$$

with $\tilde{\phi}_\theta = |g|^2 * \tilde{\phi}$.

Write $\tilde{\phi} = \tilde{\phi}_\theta + (\tilde{\phi} - \tilde{\phi}_\theta)$ and $h^\theta = -(1 - \varepsilon)\Delta - \tilde{\phi}_\theta(x)$. Let us first concentrate on $e_1 = \inf e_1(\psi)$, where

$$e_1(\psi) = \left(\psi, \sum_{i=1}^N h_i^\theta \psi \right).$$

Since $\sum h^\theta$ is a sum of single-particle operators we need only consider ψ 's which are determinants of N orthonormal single-particle functions. If $m_1, \dots, m_N$ are such, then

$$M(p, r) = \sum_{i=1}^N (m_i, \pi_{p,r} m_i)$$

has the property that $0 \leq M(p, r) \leq q$ and

$$(2\pi)^{-3} \int dp dr M(p, r) = N.$$

Therefore

$$\begin{aligned} e_1(\psi) &= (2\pi)^{-3} \iint dp dr \{ (1 - \varepsilon) p^2 - \tilde{\phi}(r) \} M(p, r) \\ &\quad - N \int |\nabla g|^2. \quad (5.21) \end{aligned}$$

The minimum of the right side of Eq. (5.21) over all M with the stated properties is given as follows:

$$M(p, r) = q \theta(\tilde{\phi}(r) - (1 - \varepsilon) p^2 - \mu)$$

for some $\mu \geq 0$. μ is the smallest μ such that $(2\pi)^{-3} \times \int M(p, r) \leq N$. Since $\tilde{\phi}$ is the TF potential [for $(1 - \varepsilon)\gamma_\rho$] we see that $\mu = \tilde{\mu}$ and

$$e_1 - D(\tilde{\rho}, \tilde{\rho}) + U \geq \tilde{E} - N \int |\nabla g|^2. \quad (5.22)$$

Next, let us consider the missing piece $e_2 = - \int (\tilde{\phi} - \tilde{\phi}_\theta) \rho_\theta$. The second piece of $\tilde{\phi}$, namely $-\tilde{\psi} = -|x|^{-1} * \rho$, has the property that $\tilde{\psi} - |g|^2 * \tilde{\psi} \geq 0$ since $\tilde{\psi}$ is superharmonic and $|g|^2$ is spherically symmetric.

Therefore we can ignore this piece in e_2 . $V - V_\theta$ is bounded above, as before, by

$$\sum z_j |x - R_j|^{-1} \theta(R - |x - R_j|).$$

For large N , $|R_i - R_j| > 2R$ and, using Hölder's inequality,

$$\begin{aligned} e_2 &\geq - \|\rho_\theta\|_{5/3} \left[8\pi R^{1/2} \sum z_j^{5/2} \right]^{2/5} \\ &\geq - \|\rho_\theta\|_{5/3} \{ (8\pi)^{2/5} R^{1/5} Z \}. \quad (5.23) \end{aligned}$$

The negative term e_2 is controlled by the εT_θ term through an inequality of Lieb and Thirring (1975 and 1976; see also Lieb, 1976):

$$T_\theta \geq L \int \rho_\theta(x)^{5/3} dx, \quad (5.24)$$

with $L = \frac{2}{3} (3\pi/2q)^{2/3}$. Furthermore, by the Schwarz inequality, $\int \rho_\theta^{4/3} \leq \{ N \int \rho_\theta^{5/3} \}^{1/2}$. If we write $\int \rho_\theta^{5/3} = X$ then $e_2 \geq -X^{2/5} D$, with $D = \{ \}$ in Eq. (5.23), and

$$\begin{aligned} e_2 + \varepsilon T_\theta - (1.68) \int \rho_\theta^{4/3} \\ \geq \min_{X>0} -DX^{2/5} + \varepsilon LX - (1.68)N^{1/2}X^{1/2} \equiv Y. \quad (5.25) \end{aligned}$$

Equation (5.22) contains $\tilde{E}$ instead of E ; we must bound the difference. Using $\tilde{\rho}$ as a trial function for E , $E \leq \tilde{E} + [\varepsilon/(1 - \varepsilon)] \tilde{K}$, where

$$\tilde{K} = [3(1 - \varepsilon)/5] \gamma_\rho \int \tilde{\rho}^{5/3}.$$

Choose $\varepsilon = Z^{-1/30}$ (this is not optimum). For large Z , $\varepsilon < \frac{1}{2}$ and it is easy to see that $\tilde{K} < (\text{const})Z^{7/3}$ for all N, Z . Thus

$$0 > \tilde{E} - E > -(\text{const})Z^{7/3}Z^{-1/30}. \quad (5.26)$$

Choose $R = Z^{-1/2}$, which is a different choice from the upper bound calculation. Then $D \approx Z^{9/10}$ and

$$Y \geq -(\text{const})Z^{7/3}Z^{-1/30}. \quad (5.27)$$

[It is easy to see that the term $-(1.68)N^{1/2}X^{1/2}$ is negligible as long as N/Z is fixed.]

Finally $-N \int |\nabla g|^2 \approx -NR^2 \approx Z^2$. Combining all these bounds, we find

$$E_N^Q - E > -(\text{const})Z^{7/3-1/30}$$

which is the desired result. ■

Clearly there is room for a great deal of improvement, for it is believed that $E_N^Q - E > 0$ as explained in Sec. V.B. But first let us turn to the correlation functions.

3. Correlation functions

In analogy with Eq. (5.14) we define

$$\begin{aligned} \rho_\theta^j(x_1, \dots, x_j) &= j! \left(\sum_j \right) \int \prod_{i=1}^j |\psi(x_i, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 \\ &\quad \times dx_{j+1} \dots dx_N. \quad (5.28) \end{aligned}$$

We wish to obtain a limit theorem for ρ_θ^j when ψ is a ground state of H_N . But there may be no ground state ($\inf \text{spec } H_N$ may not be an eigenvalue) or there may be

several. In any case, it is intuitively clear that the limit of ρ_N^t should not depend upon ψ being exactly a ground state, but only upon ψ being "nearly" a ground state.

Definition. Let $\psi_1, \psi_2, \dots$ be a sequence of normalized functions with $\psi_N \in \mathcal{H}_N$ for N particles. This sequence is called an *approximate ground state* if $|\psi_N, H_N \psi_N| - E_N^0 |a_N^{-7/3} - 0$ as $N \rightarrow \infty$. H_N always has k nuclei.

Theorem 5.2. Let $\{\psi_N\}$ be an approximate ground state with the scaling given before Eq. (5.2), and let $\rho_N^t(x)$ be given by Eq. (5.28) with ψ_N , and

$$\hat{\rho}_N^t(x_1, \dots, x_j) = a_N^{-2j} \rho_N^t(a_N^{-1/3} x_1, \dots, a_N^{-1/3} x_j).$$

Let $\rho^t(x_1, \dots, x_j) = \rho(x_1) \cdots \rho(x_j)$ with ρ being the solution to the TF problem for λ and $\{z_j^0, R_j^0\}$. (Note that $\lambda = N/Z$ is now fixed.) Then

$$\begin{aligned} & \left(\psi_N, -\sum \Delta_i \psi_N \right) a_N^{-7/3} \rightarrow \frac{2}{5} \gamma_p \int \rho(x)^{5/3} dx, \\ & \left(\psi_N, \sum V(x_i) \psi_N \right) a_N^{-7/3} \rightarrow \int \rho V, \\ & \left(\psi_N, \sum |x_i - x_j|^{-1} \psi_N \right) a_N^{-7/3} \rightarrow D(\rho, \rho). \end{aligned}$$

Moreover, $\hat{\rho}_N^t(x) \rightarrow \rho^t(x)$ in the sense that if Ω is any bounded set in $\mathbb{R}^3$ then

$$\int_{\Omega} \hat{\rho}_N^t(x) d^3x \rightarrow \int_{\Omega} \rho^t(x) d^3x.$$

If $\lambda \leq Z = \sum z_j$, the restriction that Ω be bounded can be dropped and $\hat{\rho}_N^t \rightarrow \rho^t$ in the weak L^1 sense.

Proof. The reader is referred to LS, Theorem III.5 for details. The basic idea is to consider a function $U(x_1, \dots, x_j) \in C_0^\infty(\mathbb{R}^{3j})$ and add $\alpha \int \rho^t U d^3x$ to the TF functional, $\mathcal{E}(\rho)$. On the other hand, the potential

$$\alpha a_N^{4j/3} \sum_{\substack{i_1, \dots, i_j \\ \text{unequal}}} U(a_N^{1/3} x_{i_1}, \dots, a_N^{1/3} x_{i_j})$$

is added to H_N . By the aforementioned methods the energies are shown to converge on the scale of $a_N^{7/3}$. But $\partial E / \partial \alpha|_{\alpha=0} = \int \rho^t U$. By concavity of $E(\alpha)$ the derivatives and the limits $a_N \rightarrow \infty$ can be interchanged. Thus, for all such U , $\int \hat{\rho}_N^t U \rightarrow \int \rho^t U$. ■

One of the assertions of Theorem 5.2 is that, as $N \rightarrow \infty$, correlations among any finite number of electrons disappear. *A posteriori* this is the justification for replacing the electron-electron repulsion $\sum |x_i - x_j|^{-1}$ by $D(\rho, \rho)$ in TF theory.

B. The Scott conjecture for the leading correction

We have seen that $E^{\text{TF}} = -CZ^{7/3}$ under the assumption that the nuclear coordinates R_i and charges z_j scale as $Z^{-1/3} R_j^0$ and $Z z_j^0$, $\sum z_j^0 = 1$, and $\lambda = N/Z > 0$ is fixed. C depends on λ, z_j^0, R_j^0 . What is the next correction to the energy? While this question takes us to some extent outside TF theory, we should like to mention briefly the interesting conjecture of Scott (1952) and a generalization of that conjecture. None of these conjectures have been proved.

The basic idea of Scott is that in the *Bohr atom* (no

electron repulsion) the electrons close to the nuclei *each* have an energy $\sim -Z^2$. This should also be true in some sense even with electron repulsion. Since TF theory cannot yield exactly the right energy near the singularities of V , the leading correction should be $O(Z^2)$.

The leading correction should have three properties.

- (i) It is the *same* with or without electron repulsion because the repulsive part of $\phi(x)$, namely $|x|^{-1} \star \rho$, is $O(Z^{4/3})$ for all x .
- (ii) It is independent of N/Z , provided $N/Z > 0$ and fixed. This is so because the correction comes from the core electrons whose distance from the nucleus is $O(Z^{-1})$. The number of electrons thus involved is small compared to Z .
- (iii) It should be additive over a molecule. If the correction is Dz^2 for an atom then the total leading correction should be

$$\Delta E = D \sum_{j=1}^k z_j^2 \quad (5.29)$$

and

$$E^Q = E^{\text{TF}} + \Delta E + o(Z^2). \quad (5.30)$$

Of course E^{TF} depends on whether electron repulsion is present or not, but ΔE supposedly does not change. To calculate D let us first calculate E^{TF} for an atom without repulsion. The general theory goes through as before, but now the TF equation is $\gamma \rho^{2/3} = (V - \mu)_+$, $V(x) = z/|x|$, $\int \rho = N$, and $\mu > 0$, even when $N = z$. It is found (Lieb, 1976, p. 560) that $\mu = z/R$, $R = 3\gamma(4N/\pi^2)^{2/3}/5z$, and $E_{\text{Bohr}}^{\text{TF}} = -3z^2 N^{1/3} (\pi^2/4)^{2/3} \gamma$. Using γ_p ,

$$E_{\text{Bohr}}^{\text{TF}} = -z^{7/3} (3N/z)^{1/3} (2mq^2/3/\hbar^2)/4.$$

The quantum energy is computed by adding up the Bohr levels. For each principal quantum number n , the energy is $e_n = m/2\hbar^2 n^2$ and it is qn^2 -fold degenerate. The result (Lieb, 1976) is

$$E_{\text{Bohr}}^Q = E_{\text{Bohr}}^{\text{TF}} + qz^2/8 + O(z^{5/3}), \quad (5.31)$$

thus

$$D = qz^2/8 \quad (5.32)$$

in the Scott conjecture. Scott's (1952) derivation was slightly different from the above, but his basic idea was the same.

The Scott conjecture about the energy can be supplemented by the following about the density. Let $f_{n\ell m}(z, x)$ be the normalized *bound-state* eigenfunctions for the hydrogenic atom with nuclear charge z , and define

$$\rho^H(z, x) = q \sum_{n\ell m} |f_{n\ell m}(z, x)|^2. \quad (5.33)$$

This sum converges and represents the quantum density for a Bohr atom with *infinitely many electrons*. It is being tabulated and studied by Heilmann and Lieb. It is monotone decreasing and a graphical plot of ρ^H shows that it has almost no discernible shell structure. Clearly $\rho^H(z, x) = z^3 \rho^H(1, zx)$ and is spherically symmetric. By our previous analysis of the $z \rightarrow \infty$ limit (which strictly speaking is not applicable when $N = \infty$, but which can be suitably modified)

$$z^{-2} \rho^H(z, z^{-1/3} x) \rightarrow z^{-2} \rho_{\text{Bohr}}^{\text{TF}}(z, z^{-1/3} x) \quad (5.34)$$

as $z \rightarrow \infty$. But

$$\rho_{\text{Bohr}}^{\text{TF}}(z, x) = (z/\gamma_p |x|)^{3/2}$$

when $\mu = 0$, as we have just seen. Thus

$$\rho^H(1, y) - (\gamma_p |y|)^{-3/2} \quad (5.35)$$

as $y \rightarrow \infty$. Equation (5.35) is not obvious, but it can be directly proved from (5.33).

Thus $\rho^H(z, x)$, whose scale length is z^{-1} , agrees nicely with $\rho^{\text{TF}}(z, x)$, whose scale length is $z^{-1/3}$, in the overlap region $z^{-1} \ll |x| \ll z^{-1/3}$. This is true even when electron repulsion is included in ρ^{TF} because of Theorem 2.8(a). The common value is $\rho(z, x) = (z/\gamma_p |x|)^{3/2}$. Because of this we are led to the following.

Conjecture. Suppose the sequence $\{\psi_N\} \in \mathcal{K}_N$ is an *ap-proximate ground state* for a molecule (with repulsion) in the strong sense that

$$|\langle \psi_N, H_N \psi_N \rangle - E_N^0| a_N^{-2} \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Let $\rho_N^0(x)$ be given by Eq. (5.14). Recall that $R_j = a_N^{-1/3} R_j^0$. Fix $\lambda = N/Z > 0$ and $x \neq R_j^0$, all j . Then, as $N \rightarrow \infty$,

$$a_N^{-3} \rho_N^0(a_N^{-1/3} x) - \rho^{\text{TF}}(x), \quad (5.36)$$

where ρ^{TF} is the TF density for λ, z^0, R_j^0 . On the other hand, for all fixed y ,

$$a_N^{-3} \rho_N^0(a_N^{-1/3} R_j^0 + a_N^{-1} y) - (z^0)^3 \rho^H(1, z^0 y). \quad (5.37)$$

Equation (5.36) has already been proved in Sec. V.A.

TFW Theory. It is a remarkable fact that the TFW correction, which has no strong *a priori* justification, has, as its chief effect, precisely the kind of correction (i), (ii), (iii) above predicted by Scott. If δ is chosen correctly in Eq. (2.8), even the constant D in Eq. (5.32) can be duplicated. This will be elucidated in Sec. VII. TFW theory also (accidentally?) improves TF theory in two other ways: negative ions can be supported and binding occurs.

C. A picture of a heavy atom

With the real and imagined information at our disposal we can view the energy and density profile of a heavy, neutral, nonrelativistic atom as being composed of seven regions.

(1) *The inner core.* Distances are $O(z^{-1})$ and ρ is $O(z^3)$. For large σ , the number of electrons out to $R = \sigma/z$ is $\sim \sigma^{3/2}$, while the energy $\sim z^2 \sigma^{1/2}$. If $1 \ll z r \ll z^{2/3}$, $\rho(r)$ is well approximated by $(z/\gamma_p r)^{3/2}$. $\rho(r)$ is infinity on a scale of z^2 which is the appropriate scale for the next, or TF region. The leading corrections, beyond TF theory, come from this region. None of this has been proved.

(2) *The core.* Distances are $O(z^{-1/3})$ and ρ is $O(z^2)$. TF theory is exact to leading order. The energy is $E^{\text{TF}} \sim z^{7/3}$ and almost all the electrons are in this region. This is proved.

(3) *The core mantle.* Distances are of order $\sigma z^{-1/3}$ with $\sigma \gg 1$. $\rho(r) = (3\gamma_p/\pi)^3 \gamma^{-6}$, the Sommerfeld asymptotic formula. ρ is still $O(z^2)$. This is proved.

(4) *A transition region to the outer shell.* This region may or may not exist.

(5) *The outer shell.* In the Bohr theory, $z^{1/3}$ shells are filled. The outer shell, if it can be defined, would

presumably contain $O(z^{2/3})$ electrons and each electron in the shell would "see" an effective nuclear charge of order $z^{2/3}$. This picture would give a radius unity for the last shell and an *average* density $\sim z^{2/3}$ in the shell. On the same basis the *average* electron energy would be $O(z^{2/3})$ and thus the energy in the shell would be $O(z^{4/3})$. All this is conjectural, for reliable estimates are difficult to obtain.

(6) *The surface.* Here the potential is presumably $O(1)$, and so is the energy of each electron. Chemistry takes place here.

TF theory, which is unreliable in this region, nevertheless predicts a surface radius of $O(1)$. We thank J. Morgan for this remark. His idea is that if the surface radius R_s is defined to be such that outside R_s there is one unit of electron charge, then $R_s = O(1)$ because the TF density is $\rho(r) = (3\gamma_p/\pi)^3 r^{-6}$, independent of z , for large r . Likewise, if R_0 is defined such that between R_0 and R_s there are $z^{2/3}$ electrons, then the average TF density in this "outer shell" is $z^{2/3}$ in conformity with (5). Finally, the energy needed to remove one electron is $O(1)$ as Eq. (3.11) shows. The radius of this ionized atom is also $O(1)$ as Eq. (3.13) shows.

In no sense is it being claimed that TF theory is reliable at the surface, or even that the existence of the surface, as described, is proved. We are only citing an amusing coincidence. It is quite likely that the surface radius of a large atom has a weak dependence on z .

(7) *The region of exponential falloff.* $\rho(r) \sim K \exp[-2(2me/\hbar^2)^{1/2}(r-R)]$, where e is the ionization potential, K is the density at the surface, and R is the surface radius. An upper bound for ρ of this kind has been proved by many people, of whom the first was O'Connor (1973). See also Deift, Hunziker, Simon, and Vock, 1978, and M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, R. Ahlrichs, and J. Morgan, 1980, for recent developments and bibliographies of earlier work.

The density profile of a heavy atom, as described above, is shown schematically in Fig. 2.

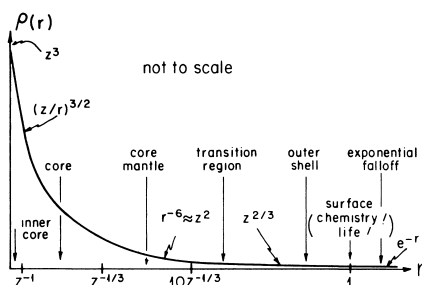


FIG. 2. Schematic plot of the electron density $\rho(r)$ in a neutral heavy atom of charge z . The inner core extends to distances of order z^{-1} ; the core to order $z^{-1/3}$; the mantle to $z^{-1/3}$ times a large number. The core and its mantle are correctly described by TF theory. The outer shell extends to distances of order z^p where p is near zero. Finally, there is the surface, and then the region of exponential falloff. The surface thickness is not shown.

VI. THOMAS-FERMI-DIRAC THEORY

The previous sections contain most of the mathematical tools for the analysis of this model; the main new mathematical idea to be introduced here will be the j model and its relation to TFD theory.

The TFD functional is

$$\mathcal{E}^{\text{TFD}}(\rho) = \int J(\rho(x)) dx - \int V\rho + D(\rho, \rho) + U, \quad (6.1)$$

and

$$J(\rho) = \frac{3}{2} \gamma \rho^{5/3} - \frac{3}{4} C_e \rho^{4/3}. \quad (6.2)$$

The term $-D = -(3C_e/4) \int \rho^{4/3}$ was suggested by Dirac (1930) to account for the “exchange” energy. The true electron repulsion I in (5.17) is expected to be less than $D(\rho, \rho)$ because the electrons are correlated. For an ideal Fermi gas at constant density, I is computed to be $D(\rho, \rho) - D$ with $C_e = (6/\pi q)^{1/3}$. There is, however, no fundamental justification for the Dirac approximation; it can even lead to unphysical results, as will be seen shortly. In particular, I is always positive but $D(\rho, \rho) - D$ can be arbitrarily negative. As remarked in (5.17), there is a lower bound of this form $D(\rho, \rho) - D$ (Lieb, 1979; Lieb and Oxford, 1981) with $3C_e/4 = 1.68$ (independent of q). In any event, it should be remembered that D is part of the Coulomb energy even though it is mathematically convenient to combine it with the kinetic energy as in Eq. (6.2).

For simplicity we assume

$$V(x) = \sum_{i=1}^N V_i(x - R_i), \quad (6.3)$$

with $V_i \in \mathcal{D}$: $V_i = |x|^{-1} * m_i$ (with m_i a non-negative measure) and $|m_i| = z_i$.

Henceforth the superscript TFD will be omitted. All quantities in this section refer to TFD, and not TF, theory, unless otherwise stated.

A. The TFD minimization problem

The function space is the same as for TF theory, namely

$$\mathcal{G} = \{\rho \mid \rho \in L^1 \cap L^{5/3}, \rho(x) \geq 0\}. \quad (6.4)$$

The energy is

$$E(\lambda) = \inf \left\{ \mathcal{E}(\rho) \mid \int \rho = \lambda, \rho \in \mathcal{G} \right\}. \quad (6.5)$$

Theorem 6.1. $E(\lambda)$ is finite, nonincreasing in λ , and

$$E(\lambda) = \inf \left\{ \mathcal{E}(\rho) \mid \int \rho \leq \lambda, \rho \in \mathcal{G} \right\}. \quad (6.6)$$

Moreover, $e(\lambda) \equiv E(\lambda) - U < 0$ when $\lambda > 0$.

Proof. Same as Proposition 2.1 and Theorem 2.3. The crucial fact to note is that $J(0) = J'(0) = 0$, which permits us to place “surplus charge density” at infinity. ■

It is not immediately obvious that $E(\lambda)$ is convex because J is not convex. The proof of convexity is complicated and will be given later (Theorem 6.9).

A second difficulty is that $E(\lambda)$ is not bounded below for all λ . This is so because J is not positive. This latter difficulty can be dealt with in the following way. Introduce

$$\mathcal{E}_\alpha(\rho) = \mathcal{E}(\rho) + \alpha \int \rho \quad (6.7)$$

with

$$\alpha = 15C_e^2(64\gamma)^{-1}. \quad (6.8)$$

This amounts to replacing J by

$$J_\alpha(\rho) = J(\rho) + \alpha \rho. \quad (6.9)$$

Note that $J_\alpha(\rho) \geq 0$ and

$$J_\alpha(\rho_0) = 0 = J'_\alpha(\rho_0) \quad (6.10)$$

for $\rho_0 = (5C_e/8\gamma)^3$. $-\alpha$ and ρ_0 are the minimum value and the minimum point of the function $J(\rho)/\rho$. Correspondingly, introduce

$$E_\alpha(\lambda) = \inf \left\{ \mathcal{E}_\alpha(\rho) \mid \int \rho = \lambda, \rho \in \mathcal{G} \right\}. \quad (6.11)$$

Theorem 6.2. $E_\alpha(\lambda)$ is nonincreasing in λ and has a lower bound, independent of λ . Moreover,

$$E_\alpha(\lambda) = E(\lambda) + \alpha \lambda \quad (6.12)$$

and

$$E_\alpha(\lambda) = \inf \left\{ \mathcal{E}_\alpha(\rho) \mid \int \rho \leq \lambda, \rho \in \mathcal{G} \right\}. \quad (6.13)$$

$e_\alpha(\lambda) \equiv E_\alpha(\lambda) - U = e(\lambda) + \alpha \lambda < 0$ when $\lambda > 0$, and $e_\alpha(\lambda) \rightarrow -\inf_e e_\alpha(\lambda) = e_\alpha(\infty)$ as $\lambda \rightarrow \infty$.

Proof. Again the proof is the same as for Proposition 2.1 and Theorem 2.3. Here, however, $J'_\alpha(0) > 0$; the fact that $J_\alpha(\rho_0) = J'_\alpha(\rho_0) = 0$ is used instead. The fact that $J_\alpha \geq 0$ is responsible for the lower bound. ■

Remark. One consequence of Theorem 6.2 is that $dE(\lambda)/d\lambda \leq -\alpha$ (if the derivative exists). Another is that when λ is large enough so that $e_\alpha(\lambda) = e_\alpha(\infty)$ then $e(\lambda) = e_\alpha(\infty) - \alpha\lambda$. As will be seen, this happens when $\lambda \geq \lambda_c = Z = \sum z_i$. Thus the graph of $e_\alpha(\lambda)$ is similar to that for $e^{\text{TF}}(\lambda)$ in Fig. 1. $e(\lambda)$ then has a negative slope, $-\alpha$, at λ_c and afterwards $e(\lambda)$ has the same constant negative slope. This is a highly unphysical feature of TFD theory which arises from the fact that one can have spatially small “clumps” of density in which $\rho = \rho_0$, arbitrarily far apart. These “clumps” have an energy approximately $-\alpha\rho_0 \cdot (\text{volume})$ and are physically nonsensical because the $-\rho^{4/3}$ term, which causes this effect, is a gross underestimate of the positive electron repulsion which it is meant to represent. There is no minimizing ρ for these “clumps” because for no ρ is the energy exactly $-\alpha\rho_0 \cdot (\text{volume})$. The “inf” in Eq. (6.5) is crucial.

B. The j model

Now we must deal with the fact that J_α is not convex. To this end we follow Benguria (1979), who introduced the “convexified” j model. With its aid, Benguria was the first to place the TFD theory on a rigorous basis for a certain class of amenable potentials in Eq. (6.3), which is defined in Sec. VI.C. This class includes the point nuclei. It will turn out that the j model also permits us to analyze TFD theory for all potentials, not just the amenable class. However, for nonamenable potentials, the analysis is complicated and the final result has an unexpected feature, namely, that a minimizing ρ for E may not exist, even if $\lambda < \lambda_c$. The j model is explored in

detail here because, as will be seen in Sec. VI.C, its energy is the same as $E_\alpha(\lambda)$ for the TFD model. Moreover, for amenable potentials the density ρ of the two models is also the same.

Definition.

$$j(\rho) = J_\alpha(\rho), \quad \rho \geq \rho_0 = (5C_e/8\gamma)^3 \\ = 0, \quad 0 \leq \rho \leq \rho_0. \quad (6.14)$$

The derivative of this convex function is given in Eq. (3.14). $\mathcal{E}_j(\rho)$ is given by Eq. (6.1) with J replaced by j . $E_j(\lambda)$ is defined by Eq. (6.5) with $\mathcal{E}$ replaced by $\mathcal{E}_j$.

By the methods of Secs. II and III the j model has many of the same properties as TF theory.

Theorem 6.3. *If V is given by Eq. (6.3) and if $\mathcal{E}$ is replaced by $\mathcal{E}_j$, E by E_j , and e by $e_j = E_j - U$, then the following results of TF theory hold for the j model (they also hold for TF theory, of course, with this V) (Ignore any mention of TFD and TFDW theory in the cited theorems.):*

Propositions 2.1 and 2.2; Theorems 2.3 and 2.4; the definition of λ_c ; Theorem 2.5; Theorem 2.6 [with Eq. (2.18) replaced by (3.2)]; Theorem 2.7; Theorem 2.12 (for a point nucleus); Theorem 2.13 without the γ dependence (for point nuclei); the last two equations in this theorem have an obvious generalization for non-point nuclei.); Theorem 2.14 (for point nuclei) is changed to (a) $5K/3 = A - 2R - \mu\lambda - \alpha\lambda + 4D/3$, (b) $2K = A - R + D$ for an atom, with $D = (3C_e/4)\int \rho^{4/3}$ (note that Theorem 3.19 must be used in the proof); Equation (2.22); Theorem 3.2; Theorems 3.4, 3.5 [Benguria (1979) has shown that if W is the potential of point nuclei then $\phi' - \phi \in H^2$ away from S_W]; Corollaries 3.7, 3.8, 3.9, and 3.10 (note, in particular, that $\phi^{TF} > \phi^{model}$ for fixed μ); Lemma 3.11; Theorem 3.12, Corollaries 3.14 and 3.17; Theorem 3.18 (i.e., $\lambda_c = Z$); Equation (3.6); Sec. III.B; Theorem 3.23 (but note that equality can occur. See remark at the end of Sec. IV.C).

Remarks. (i) Theorem 2.8(a) holds in the sense that $\rho(x) \approx (z/\gamma)^{3/2} |x - R_j|^{-3/2}$ near R_j .

(ii) There is no simple scaling for the j model, as in Eq. (2.24) for TF theory.

(iii) We emphasize that a minimizing ρ exists if and only if $\lambda \leq Z$. This ρ is unique and satisfies the Thomas-Fermi-Dirac equation (3.2).

(iv) *Question.* Under what conditions do the conclusions of Corollaries 3.13, 3.15, and 3.16 and Theorem 3.21 hold for the j model? *Question.* To what extent do the results of Sec. IV carry over to the j model?

(v) To prove the analogue of Eq. (2.15), Mazur's theorem can be used, as in the proof of Proposition 3.24.

There are some useful additional facts about the j model not mentioned in Theorem 6.3.

Theorem 6.4. *If C_e increases then (i) $\phi(x) - \mu(\lambda)$ decreases and $\mu(\lambda)$ increases, for fixed λ ; (ii) $\phi(x)$ decreases for fixed μ .*

Proof. By Corollary 3.10, since $j'(\rho)$ decreases with C_e for fixed ρ . ■

Theorem 6.5. *For all λ , $E^{TF}(\lambda) > E_j(\lambda) > E(\lambda)$ since $J(\rho) < j(\rho) < 3\gamma\rho^{5/3}/5$. On the other hand, suppose V is*

the potential of k point nuclei as in Eq. (2.1). Then for $\lambda \leq \lambda_c = Z$,

$$E^{TF}(\lambda) \leq E_j(\lambda) - \alpha\lambda + (3C_e/4)\lambda^{1/2} \{ (5\epsilon_1/2\gamma)Z^{7/3} \}^{1/2} \\ + 27C_e^2\lambda/(10\gamma), \quad (6.15)$$

where $-\epsilon_1$ is the TF energy for a neutral atom with $z=1$ [see Eq. (7.15)].

Remarks. (i) When $\lambda > Z$ then $(E^{TF} - E_j)(\lambda) = (E^{TF} - E_j)(Z)$.

(ii) Clearly Eq. (6.15) can be improved. But it does show that the effect of the Dirac term is to decrease the energy by $O(Z^{5/3})$ for large Z . (Note: by Theorem 6.8, $E_j - \alpha\lambda = E$.)

Proof. Let ρ be the minimizing density for E_j , and ρ^{TF} that for E^{TF} . Use ρ as a trial function for E^{TF} . Noting that $\rho(x) \neq 0, \rho_0$ a.e. (Theorem 3.19), we have $E^{TF} \leq \mathcal{E}^{TF}(\rho) = E_j - \alpha\lambda + (3C_e/4)\int \rho^{4/3}$. By Theorem 3.19, $\gamma\rho^{2/3} - C_e\rho^{1/3} + \alpha = \phi - \mu$ when $\rho > 0$. But by Corollary 3.10, $\phi - \mu \leq \phi^{TF} - \mu^{TF} \leq \gamma(\rho^{TF})^{2/3}$. Thus $\rho^{2/3} \leq (\rho^{TF})^{2/3} + C_e/\gamma\rho^{1/3}$. Squaring this and using $\int (\rho^{TF})^{2/3}\rho^{1/3} \leq \lambda$ (Hölder), and $\int \rho^{2/3} \leq \lambda(\rho_0)^{-1/3}$, and $\int (\rho^{TF})^{4/3} \leq [\lambda \int (\rho^{TF})^{5/3}]^{2/2}$, we obtain Eq. (6.15), but with $5K^{11}/3\gamma$ in place of $\{ \}$. By Theorem 2.14(a), $2K^{11}/3 < -e^{11}$, and by the remark preceding Eq. (4.5), $e^{11} > e^{11}$ (all nuclei at one point). ■

The next theorem states that ρ always has compact support, even when $\lambda = \lambda_c$. When $\lambda < \lambda_c$ this is also true in TF theory (Lemma 3.11). The proof we give seems unnecessarily complicated; a simpler one must be possible.

Theorem 6.6. *Suppose $V = |x|^{-1} * m \in \mathcal{D}$, with m a non-negative measure of compact support and $\int dm = Z$. Let ρ be the minimizing j model density for $\lambda \leq \lambda_c = Z$. Then ρ has compact support. Moreover, suppose $\text{supp}(m) \subset B_R$, the ball of radius R centered at 0. Then $\text{supp}(\rho) \subset B_r$ for some r depending on R and Z , but independent of λ . $r \leq 2R + tZ(\rho_0 R^2)^{-1}$ for some universal constant t , independent of all parameters.*

Proof. The strategy is to construct a function f such that $\text{supp}(\rho) \subset \text{supp}(f)$. Let $S_R = \partial B_R$ be the sphere of radius R . There exists a function (surface charge distribution) σ on S_{2R} such that V_σ , the potential of σ outside B_{2R} , is V , i.e.,

$$V(x) = V_\sigma(x) \equiv 4\pi(2R)^2 \int d\Omega \sigma(\Omega) |x - 2R\Omega|^{-1}$$

for $|x| > 2R$, where Ω denotes a point on S_1 and $\int d\Omega = 1$. It is easy to see that σ is a bounded, continuous function since $\text{supp}(m) \subset B_R$, and $|\sigma(\Omega)| \leq sZR^{-2}$ for some universal constant s . Let $\Sigma(\Omega) = -\sigma(\Omega) + sZR^{-2} \geq 0$, and let

$$dM(x) = dm(x) + \Sigma(x/2R)\delta(|x| - 2R)dx.$$

If $V_M = |x|^{-1} * M$, we see that $V_M(x) \geq V(x)$, all x , and $V_M(x) = Q|x|^{-1}$ for $|x| > 2R$, with $0 \leq Q \leq (1 + 16\pi s)Z$. $V_M - V$ is a bounded function in $\mathcal{D}$. Now let $f(x) = \rho_0$ for $2R < |x| \leq r$ and $f(x) = 0$ otherwise, with $\int f = Q$. Let $\psi = |x|^{-1} * \rho$, $g = |x|^{-1} * f$, $\phi = V - \psi$, $h = V_M - g$ where ρ satisfies Eq. (3.2). We claim $u(x) \equiv h(x) - \phi(x) + \mu \geq 0$. If not, let $B = \{x | u(x) < 0\}$. B is open since $h - \phi$ is con-

tinuous. $-\Delta u/4\pi \geq -f + \rho$. But $f \leq \rho$ on B since, for $|x| \leq 2R$, $f(x) = 0$; for $|x| > 2R$, $\rho(x)$ is either 0 or $> \rho_0$ a.e. by Theorem 3.19. If $\rho(x) = 0$ then $\phi - \mu \leq 0$ by (3.2) and $h(x)$ is clearly ≥ 0 , so $x \notin B$; if $\rho(x) \geq \rho_0$, $\rho - f \geq 0$. Thus u is superharmonic on B and, since $u = 0$ on ∂B and $u = \mu \geq 0$ at infinity, B is empty. Now consider $A = \{x | r < |x|\}$. In A , $-\Delta u = 4\pi\rho \geq 0$ and $u \geq 0$ on ∂A and at infinity. Therefore either (i) $\rho = 0$ a.e. in A or (ii) $u > 0$ everywhere in A . In case (ii), $\phi < \mu$ in A because $h = 0$ in A . But then, by Eq. (3.2), $\rho = 0$ a.e. in A . The bound for r is obtained by $4\pi(r - 2R)(2R)^2 < (4\pi/3)[r^3 - (2R)^3] = Q/\rho_0$. ■

Remark. For an atom with nucleus located at the origin, R can be chosen to be any positive number. If the inequality for r is minimized then we find $\rho(x) = 0$ for $|x| > 3(tZ/\rho_0)^{1/3}$.

Theorem 6.7. Suppose $V \in \mathfrak{D}$ and $\rho \in \mathfrak{C}$ are such that the second line of Eq. (3.2) holds with $\mu = 0$, in the sense that $\phi_\rho(x) \leq 0$ a.e. when $\rho(x) = 0$ and $\rho(x) = 0$ a.e. when $\phi_\rho(x) < 0$. Let A be the complement of the support of ρ . Then $\phi \equiv 0$ on $\bar{A}$, the closure of A .

Remark. Theorem 6.7 does not mention j . However, the theorem is meaningful only if $\text{supp}(\rho)$ is not all of $\mathbb{R}^3$. This does not happen in TF theory when $\mu = 0$, but it does happen for the j model if the hypothesis of Theorem 6.6 holds. The significance of Theorem 6.7 is that there is *total shielding* in TFD theory. This is in contrast to TF theory, where there is *under-screening* in the neutral case in the sense that the potential falls off only with a power law. One consequence of Theorem 6.7 is that two or more molecules, each of fixed shape, do not interact when their supports are disjoint. See the remark at the end of Sec. IV.

Proof. Let $B = \{x | \phi(x) < 0\}$. Clearly the singularities of V are not in B , so B is open. On B , $\Delta\phi \leq 0$ since $\rho = 0$ a.e. in B . But $\phi = 0$ on ∂B and at infinity so B is empty. Therefore $\phi \geq 0$ everywhere. Let $D = \{x | \rho(x) = 0\}$. $\phi \leq 0$ a.e. on D . Since $A \subset D$ is open, and ϕ is continuous on $\{x | \phi(x) < 1\}$, $\phi \equiv 0$ on A , and hence on $\bar{A}$. ■

C. The relation of the j model to TFD theory

We shall show that the energy of the j model is exactly $E_\alpha(\lambda) = E(\lambda) + \alpha\lambda$ for the TFD problem. Thus all the facts about the energy in Theorems 6.3 and 6.5 hold for TFD theory. However, the densities may be different!

Let us start with the simplest case studied by Benguria (1979).

Definition. A (non-negative) measure m is said to be *amenable* if

$$dm(x) = \sum_{i=1}^k z_i \delta(x - R_i) dx + g(x) dx$$

with $z_i > 0$ and g satisfies: (i) $g \geq 0$, (ii) $g \in L^\infty_{\text{loc}}$, (iii) If $A = \{x | g(x) = 0\}$ and $\sim A$ is its complement then $\mathbb{R}^3 \setminus [\text{Interior}(A) \cup \text{Interior}(\sim A)]$ has zero Lebesgue measure, (iv) $g(x) \geq \rho_0$ for $x \notin A$. (v) $V = |x|^{-1} * m \in \mathfrak{D}$. m is *strongly amenable* if $g(x) \geq \rho_0$ for $x \notin A$.

Remark. This amenable class is more restrictive than

necessary for Theorem 6.8. Technicalities aside, (iv) is the crucial point. δ measures (corresponding to point nuclei) are strongly amenable.

Theorem 6.8. Suppose that in Eq. (6.1) $V = |x|^{-1} * m$ and m is amenable. Then $E_\alpha(\lambda) = E(\lambda) + \alpha\lambda$ for the TFD problem equals $E_j(\lambda)$ for the j model. Moreover, there is a minimizing ρ for the TFD problem if and only if $\lambda \leq \lambda_c = Z = \int dm$. This ρ is unique and is the same as the ρ for the j model. It satisfies Eq. (3.2).

Proof. Clearly $E_\alpha \geq E_j$ since $J_\alpha(\rho) \geq j(\rho)$. First suppose $\lambda \leq \lambda_c = Z$ and let ρ be the unique minimum for the j problem. By Theorem 3.19, $\rho(x) \notin (0, \rho_0)$ so $E_\alpha(\lambda) \leq \mathcal{E}_\alpha(\rho) = \mathcal{E}_j(\rho) = E_j(\lambda)$. Thus $E_\alpha(\lambda) = E_j(\lambda)$. Let ρ satisfy $\int \rho = \lambda$ and $\mathcal{E}_\alpha(\rho) = E_\alpha(\lambda)$. Then since $\mathcal{E}_\alpha(\rho) \geq \mathcal{E}_j(\rho) \geq E_j(\lambda)$ we conclude that ρ minimizes $\mathcal{E}_j(\rho)$. But there is only one such ρ . Next, suppose $\lambda > \lambda_c$. Then $E_j(\lambda) = E_j(\lambda_c) = E_\alpha(\lambda_c)$. But $E_\alpha(\lambda) \leq E_\alpha(\lambda_c)$ by Theorem 6.2, and $E_\alpha(\lambda) \geq E_j(\lambda)$. Hence $E_\alpha(\lambda) = E_j(\lambda)$. By the above argument, any minimizing ρ for $\mathcal{E}_\alpha$ would have to minimize $\mathcal{E}_j$, but no such ρ exists. ■

Remark. By Theorem 3.19, $\rho(x) \notin (0, \rho_0)$ a.e. if m is amenable, and $\rho(x) \neq \rho_0$ a.e. if m is strongly amenable. If ρ is merely amenable, $\rho(x)$ can be ρ_0 with positive measure. An example is $dm(x) = \rho_0 B_R(x) dx$, with B_R being the characteristic function of a ball of radius R centered at 0. Then $\rho_\lambda(x) = \rho_0 B_r(x)$ with $4\pi\rho_0 r^3/3 = \lambda$ for $\lambda \leq \lambda_c = 4\pi\rho_0 R^3/3$. This ρ_λ is easily seen to satisfy Eq. (3.2).

If m is not amenable the situation is much more complicated, but more amusing mathematically. First let us consider the energy.

Theorem 6.9. If $V = |x|^{-1} * m \in \mathfrak{D}$, then $E_\alpha(\lambda) = E_j(\lambda)$ for all λ . In particular, $\lambda_c = Z = \int dm$ and E_α is convex in λ . If there is a minimizing ρ for $E(\lambda)$, it is unique and it is the ρ for the j model.

A number of lemmas are needed for the proof.

Lemma 6.10. Let $A \subset \mathbb{R}^3$ be a measurable set and let ρ be a function in L^1 with $0 \leq \rho(x) \leq 1$ for $x \in A$, and $\rho(x) = 0$ for $x \notin A$. (This implies $\rho \in L^p$, all p .) Then there exists a sequence of functions $f^n \in L^1$ such that (i) $f^n \rightarrow \rho$ weakly in every L^p with $1 < p < \infty$; (ii) f^n is the characteristic function of some measurable set $F^n \subset A$; (iii) $\int f^n = \int \rho$.

Proof. For $\delta > 0$ and $y \in \mathbb{Z}^3$ let $B(\delta, y) = \{x \in \mathbb{R}^3 | -\delta/2 < x^i - \delta y^i \leq \delta/2\}$ be the elementary cubes of side δ . Let $C(\delta, y) = A \cap B(\delta, y)$. Partition $C(\delta, y)$ into two disjoint measurable sets, C^+ and C^- , such that $|C^+(\delta, y)| = \int B(\delta, y) \rho$. Let $F^\delta = \bigcup_{y \in \mathbb{Z}^3} C^+(\delta, y)$, and let f^δ be the characteristic function of F^δ . Clearly $f^\delta \in L^p$ with norm $(\int \rho)^{1/p}$, and f^δ satisfies (ii) and (iii). Let $1/p + 1/q = 1$. Since C_0 , the continuous functions of compact support, are dense in L^q , and $\|f^\delta\|_q^q = \text{constant}$, it suffices to prove that $\int (f^\delta, g) = \int g(f^\delta - \rho) \rightarrow 0$ as $\delta \rightarrow 0$ for every $g \in C_0$. But g is uniformly continuous, so for any $\epsilon > 0$, $|g(x) - g(\delta y)| < \epsilon$ (uniformly) for $x \in B(\delta, y)$ when δ is small enough. Since

$$\int_{B(\delta, y)} (f^\delta - \rho) = 0, \quad |\int (f^\delta, g)| < 2\epsilon \int \rho. \quad \blacksquare$$

Lemma 6.11. Suppose $\rho \in \mathfrak{g}$ and

$$A(\rho) = \{x \mid 0 < \rho(x) < \rho_0\} \quad (6.16)$$

has positive measure. Then there exists $\tilde{\rho} \in \mathcal{G}$ satisfying (i) $\mathcal{G}(\tilde{\rho}) < \mathcal{G}(\rho)$; (ii) $A(\tilde{\rho})$ is empty; (iii) $\tilde{\rho}(x) = \rho(x)$ if $x \notin A(\rho)$; (iv) $\int \tilde{\rho} = \int \rho$.

Proof. Apply Lemma 6.10 to the function $h(x) = \rho(x)/\rho_0$ if $x \in A(\rho)$, $h(x) = 0$ otherwise. Let $\tilde{\rho}^n(x) = \rho_0 \int^n(x)$ if $x \in A(\rho)$, $\tilde{\rho}^n(x) = \rho(x)$ otherwise. Then $\int \tilde{\rho}^n = \int \rho$ and $\int J_\alpha(\tilde{\rho}^n) - \int J_\alpha(\rho) = -K$ with $K > 0$ and independent of n [since $J_\alpha(t) > 0$ for $0 < t < \rho_0$, and $0 = J_\alpha(0) = J_\alpha(\rho_0)$]. Now, as in the proof of Theorem 2.4, $\int V \tilde{\rho}^n \rightarrow \int V \rho$. $\lim D(\tilde{\rho}^n, \tilde{\rho}^n) = D(\rho, \rho)$ (easy proof). Hence for any $\varepsilon > 0$ there is some n such that $\int V \tilde{\rho}^n > \int V \rho - \varepsilon$ and $D(\tilde{\rho}^n, \tilde{\rho}^n) < D(\rho, \rho) + \varepsilon$. ■

The following is a corollary of Lemma 6.11.

Theorem 6.12. (a) If $\rho \in \mathcal{G}$ minimizes $\mathcal{G}(\rho)$ on $\int \rho = \lambda$, then measure $\{A(\rho)\} = 0$. (b) Even if there is no minimizing ρ , a minimizing sequence for $E(\lambda)$ with $\int \rho = \lambda$ can be chosen such that $A(\rho^n)$ is empty for all n .

Theorem 6.13. Let ρ minimize $\mathcal{G}_j$ for $\int \rho = \lambda$. $A(\rho)$ may not be empty, but for any $\varepsilon > 0$ there exists a $\tilde{\rho}$ such that $\int \tilde{\rho} = \lambda$, $A(\tilde{\rho})$ is empty, and $\mathcal{G}_j(\tilde{\rho}) < E_j(\lambda) + \varepsilon$.

Proof. Again, use Lemma 6.10 and mimic the proof of Theorem 6.11. ■

Proof of Theorem 6.9. That $E_\alpha(\lambda) = E_j(\lambda)$ follows from Theorem 6.13 and an imitation of the argument in Theorem 6.8. By Theorem 6.12 any minimizing ρ has $A(\rho)$ empty and thus minimizes $\mathcal{G}_j(\rho)$. There can be only one such ρ , since the minimizing ρ for $\mathcal{G}_j$ is unique. ■

In summary: $E(\lambda) = E_j(\lambda) - \alpha\lambda$ always, but a minimizing ρ may or may not exist for the TFD problem. It exists if and only if the minimizing ρ for the j model (which always exists when $\lambda \leq \lambda_c$) satisfies $\rho(x) \notin (0, \rho_0)$ a.e. A sufficient condition on V for this to occur is that V be amenable; a necessary condition seems to be difficult to find. The example of Sec. III.B illustrates the nonexistence phenomenon: If $dm(x) = g(x)dx$ with $g(x) \in (0, \rho_0)$ and $g \in L^1$, then $\rho_j(x) = g(x)$ in the neutral case, $\lambda = \lambda_c = \int g$. But $g(x)$ does not minimize $\mathcal{G}(\rho)$. A sequence of minimizing ρ 's for $\mathcal{G}(\rho)$ are functions which on the average locally imitate g but which oscillate rapidly between the two values 0 and ρ_0 .

VII. THOMAS-FERMI-VON WEIZSÄCKER THEORY

TFW theory was originally suggested by von Weizsäcker (1935) with $\delta = \delta_1 = \hbar^2/2m$ in Eq. (2.8). There is no fundamental justification for TFW theory in the sense that there is for TF theory, i.e., there is no theorem that the correction to the energy or density caused by the Weizsäcker term with $\delta = \delta_1$ agrees with the real quantum problem. We shall see, however, that if $\delta = (0.186)\delta_1$ then the energy correction proposed by Scott (Sec. V.B), but not the density correction conjectured in Eq. (5.37), is realized in TFW theory. If Scott is correct then *a posteriori* TFW theory has some fundamental meaning for atoms and molecules.

We were able to make a great deal of progress with TF and TFD theories essentially because of the point-

wise relation between $\phi(x)$ and $\rho(x)$. In TFW theory this relation is lost, and therefore TFW theory is much more difficult mathematically. However, the physical consequences of TFW theory are much richer and qualitatively more nearly parallel the physics of real atoms and molecules. In addition to the above mentioned Z^2 energy correction, TFW theory remedies three defects of TF (and TFD) theory:

- (i) ρ will be finite at the nuclei.
- (ii) binding of atoms occurs and negative ions are stable (i.e., $\lambda_c > Z$). These two facts are closely related.
- (iii) ρ has exponential falloff if $\lambda < \lambda_c$, e.g., for neutral atoms and molecules.

The theory presented here was begun by Benguria (1979) and then further developed by Benguria, Brezis, and Lieb (1981) (BBL), to which we shall refer for technical details. Some newer results will also be given, especially that $\lambda_c > Z$ for molecules, the Z^2 correction to the energy (Sec. VII.D), and the binding of equal atoms. Many interesting problems are still open, however.

The TFW energy functional (see Note (iv) below) is

$$\mathcal{G}(\rho) = A \int [\nabla \rho^{1/2}(x)]^2 dx + (\gamma/\rho) \int \rho(x)^{\gamma} dx - \int V(x) \rho(x) dx + D(\rho, \rho) + U. \quad (7.1)$$

This agrees with Eq. (2.8) in units in which $\hbar^2/2m = 1$. A closely related functional, obtained by writing $\psi^2 = \rho$, is

$$\mathcal{G}'(\psi) = A \int (\nabla \psi)^2 + (\gamma/\rho) \int \psi^{2\gamma} - \int V \psi^2 + D(\psi^2, \psi^2) + U. \quad (7.2)$$

Note. (i) In this section all quantities refer to TFW theory unless otherwise stated.

(ii) Equation (7.1) is defined for $\rho(x) \geq 0$ while in Eq. (7.2) $\psi(x)$ only has to be real.

(iii) As in Sec. VI, we shall assume for simplicity that V is given by Eq. (6.3) *et seq.* Later on a slightly stronger hypothesis (7.12) will be used.

(iv) $p > 1$ is a parameter; it will not be indicated explicitly unless necessary. Recall that E^{TF} was finite for point nuclei only if $p > \frac{5}{3}$. E is finite in TFW theory for all $p > 0$. We need $p > 1$ for Theorems 7.1 and 7.2, among other reasons. Even though we are interested in $p = \frac{5}{3}$, we allow p to be arbitrary because the dependence on p is interesting. It will turn out that $p = \frac{5}{3}$, the case of physical interest, is special—at least it is so for the proof that $\lambda_c > Z$.

A. The TFW minimization problem

The function space is

$$G_p' = \{ \psi \mid \nabla \psi \in L^2, \psi \in L^6 \cap L^{2p}, D(\psi^2, \psi^2) < \infty \}. \quad (7.3)$$

We say $\rho \in G_p$ if $\rho(x) \geq 0$ and $\rho^{1/2} \in G_p'$. G_p' contains

$$F_p' = G_p' \cap L^2 = \{ \psi \mid \nabla \psi \in L^2, \psi \in L^6 \cap L^{2p} \cap L^2 \}. \quad (7.4)$$

Even though we are interested in $\rho \in L^1$ (or $\psi \in L^2$), the larger space G_p' is used for technical reasons in order

to prove that $\lambda_c < \infty$; in other words, we shall eventually find that all ρ 's of interest are in F_ρ (defined analogously to G_ρ).

Remark. By Sobolev's inequality, $\|\nabla\psi\|_2 \geq L\|\psi\|_6$ with $L = 3^{1/2}(\pi/2)^{2/3}$ (cf. Lieb, 1976) when $\psi \in L^q$, $1 \leq q < \infty$. Thus when $\psi \in F'_\rho$ the restrictions that $\psi \in L^6$ and $\psi \in L^{2p}$ (if $p \leq 3$) are unnecessary. In short, $F'_\rho = H^1 \cap L^{2p}$, where $H^1 = \{\psi | \nabla\psi \text{ and } \psi \in L^2\}$.

As usual,

$$\begin{aligned} E(\lambda) &= \inf \left\{ \mathcal{E}(\rho) \mid \rho \in F_\rho, \int \rho = \lambda \right\}, \\ E'(\lambda) &= \inf \left\{ \mathcal{E}'(\psi) \mid \psi \in F'_\rho, \int \psi^2 = \lambda \right\}. \end{aligned} \quad (7.5)$$

Theorem 7.1. $\mathcal{E}(\rho)$ is strictly convex in ρ .

Proof. The only term that has to be checked is $\int (\nabla\rho^{1/2})^2$. If $\psi_i = \rho_i^{1/2}$, $i = 1, 2$, and $\psi = (\sum \alpha_i^2 \rho_i)^{1/2}$, $\sum \alpha_i^2 = 1$, then

$$\psi \nabla \psi = \sum (\alpha_i \psi_i)(\alpha_i \nabla \psi_i)$$

and

$$(\psi \nabla \psi)^2 \leq \left(\sum \alpha_i^2 \psi_i^2 \right) \left(\sum \alpha_i^2 (\nabla \psi_i)^2 \right).$$

Assuming $\psi(x) > 0$ everywhere, we are done. Otherwise, the result follows by approximation. ■

Remark. $\mathcal{E}'(\psi)$ is not convex in ψ because of the $-\int V\psi^2$ term.

Theorem 7.2. For all finite λ

$$(i) E(\lambda) = E'(\lambda);$$

$$(ii) E(\lambda) = \inf \left\{ \mathcal{E}(\rho) \mid \rho \in F_\rho, \int \rho \leq \lambda \right\}$$

and similarly for $E'(\lambda)$;

$$(iii) E(\lambda) \text{ is convex and monotone nonincreasing in } \lambda.$$

Proof. (i) Given ρ , we can always construct $\psi = \rho^{1/2}$, so $E'(\lambda) \leq E(\lambda)$. Conversely, given ψ let $f = |\psi|$. But $\nabla f = (\nabla\psi)(\text{sgn}\psi)$ so $\int (\nabla f)^2 = \int (\nabla\psi)^2$. Thus $\mathcal{E}'(f) = \mathcal{E}'(\psi)$. Choosing $\rho = f^2$, $E(\lambda) \leq E'(\lambda)$. (ii) As before, "excess charge" can be put at "infinity." Here $p > 1$ is essential. (iii) $\mathcal{E}(\rho)$ is convex so $E(\lambda)$ is convex. Monotonicity is implied by (ii). ■

Remark. (i) relates the two problems defined by Eqs. (7.1) and (7.2). To obtain the convexity (iii), $\mathcal{E}$ and Theorem 7.1 were used. We shall use $\mathcal{E}'$ to obtain the existence of a minimum, and then the TFW equation for this minimum.

Lemma 7.3. Let $V = |x|^{-1} * m$, with m a measure and $|m| = Z < \infty$. Let $\rho(x) \geq 0$. Then there exists a constant C independent of m and ρ such that for every $\varepsilon > 0$

$$\int V\rho \leq \varepsilon Z \|\rho\|_3 + Z\varepsilon^{-1/2} CD(\rho, \rho)^{1/2}.$$

Proof. By regarding R^3 as the union of balls of unit radius centered on the points of $(\frac{1}{2})Z^3$ it suffices to assume $\text{supp}(m) \subset B_1$, where $B_R = \{x \mid |x| \leq R\}$ and χ_R is the characteristic function of B_R . In the following, irrelevant constants will be suppressed. Write $V = V_- + V_+$.

where $V_- = V\chi_2$ and similarly for ρ . First consider $I_- = \int V_- \rho_-$. By Young's inequality (and writing $|x-y|^{-1} = |x-y|^{-1}\chi_2(x-y)$ for $x, y \in B_2$) $\|V_-\|_2 \leq Z$. Thus

$$\begin{aligned} I_- &\leq Z \|\rho_-\|_2 \leq Z \|\rho\|_2^{3/4} \|\rho_-\|_1^{1/4} \\ &\leq Z \{ \|\rho\|_3 + \|\rho_-\|_1 \}. \end{aligned}$$

But

$$D(\rho, \rho) \geq D(\rho_-, \rho_-) \geq \|\rho_-\|_2^2$$

since $|x-y|^{-1} > \frac{1}{8}$ in B_2 . Now, outside B_2 , $V(x) < 2Z/|x| \equiv W(x)$. Let $Q = q\chi_2$ be the constant charge distribution such that $|x|^{-1} * Q = W(x)$ outside B_2 . Then $I_+ \leq 2D(Q, \rho_+) \leq 2D(Q, Q)^{1/2} D(\rho_+, \rho_+)^{1/2}$. But $D(Q, Q) = Z^2$. Therefore (on the whole of R^3) $I \leq Z \|\rho\|_3 + ZCD(\rho, \rho)^{1/2}$. Now replace $\rho(x)$ by $\varepsilon^3 \rho(\varepsilon x)$ and $dm(x)$ by $\varepsilon^3 dm(\varepsilon x)$. Then $\varepsilon I = I_\varepsilon \leq Z\varepsilon^2 \|\rho\|_3 + ZC\varepsilon^{1/2} D(\rho, \rho)$. ■

From the Sobolev inequality, the following is obtained:

Corollary 7.4. There are constants a and $b > 0$ such that for every $\psi \in G'_\rho$

$$\mathcal{E}'(\psi) \geq a[\|\nabla\psi\|_2^2 + \|\rho\|_3^p + \|\rho\|_3 + D(\rho, \rho)] + U - b, \quad (7.6)$$

with $\rho = \psi^2$. In particular, $\mathcal{E}'$ is bounded below on G'_ρ and $E(\lambda)$ is bounded below.

It is obvious that $E(\lambda) \geq E^{\text{TF}}(\lambda)$, with the same ρ . If $p > \frac{3}{2}$, $E^{\text{TF}}(\lambda)$ is finite for point nuclei. The following illustrates the sort of lower bound for $E(\lambda)$ that can be obtained with the Sobolev inequality.

Theorem 7.5. Let $p = \frac{5}{3}$. Let $E^{\text{TFW}}(A, \gamma, \lambda)$ denote the TFW energy and $E^{\text{TF}}(\gamma, \lambda)$ denote the TF energy. Let $\tilde{L} = 9.578$. Then

$$E^{\text{TFW}}(A, \gamma, \lambda) \geq E^{\text{TF}}(\gamma + A\tilde{L}\lambda^{-2/3}, \lambda). \quad (7.7)$$

In particular, for an atom with a point nucleus, $E^{\text{TF}}(\gamma, \lambda) \sim \gamma^{-1}$ whence, for an atom,

$$E^{\text{TFW}}(A, \gamma, \lambda) \geq \gamma(\gamma + A\tilde{L}\lambda^{-2/3})^{-1} E^{\text{TF}}(\gamma, \lambda). \quad (7.8)$$

Proof.

$$\int |\nabla\psi|^2 \geq \tilde{L} \left(\int |\psi|^{10/3} \right) \left(\int |\psi|^2 \right)^{-2/3}.$$

See Lieb, 1976. ■

Remark. The right side of Eq. (7.8) has two properties: (i) Its slope at $\lambda = 0$ is finite. (ii) It is strictly monotone decreasing for all λ . To some extent, E^{TFW} will be seen to mimic this: E^{TFW} has a finite slope at $\lambda = 0$ and is strictly decreasing up to $\lambda_c > Z$.

Theorem 7.6. (i) $\mathcal{E}'(\psi)$ has a minimum on the set $\psi \in F'_\rho$ and $\int \psi^2 \leq \lambda$.

(ii) $\mathcal{E}'(\psi)$ has a minimum on G'_ρ .

(iii) The same is true for $\mathcal{E}(\rho)$ on F_ρ ($\int \rho \leq \lambda$) and G_ρ . Furthermore ρ and ψ are related by $\rho(x) = \psi(x)^2$. The minimizing ρ is unique.

Proof. The proof we give is different from the proof of Theorem 2.4 because Fatou's lemma will be used, as stated in the remark after Theorem 2.4. Let ψ^n be a minimizing sequence. By Corollary 7.4 all quantities in Eq. (7.6) are bounded. By passing to a subsequence we can demand, by the Banach-Alaoglu theorem, that $\nabla\psi^n \rightharpoonup f$ weakly in L^2 and $\rho^n \rightharpoonup \rho$ weakly in L^3 and in L^p [where

$\rho^n = (\psi^n)^2$. Furthermore, for any bounded ball B , $\psi \in L^2(B)$ since $\psi \in L^6(B)$. Moreover, $H^1(B)$ is relatively (norm) compact in $L^2(B)$. Thus, by passing to a further subsequence, we may assume $\psi^n \rightarrow \psi$ strongly in $L^2(B)$ for every B and pointwise a.e. Then it is clear that $\rho = \psi^2$ and $f = \nabla \psi$. As before, $\liminf \|\nabla \psi^n\|_2 \geq \|\nabla \psi\|_2$. From the pointwise convergence and Fatou's lemma, $\liminf D(\rho^n, \rho^n) \geq D(\rho, \rho)$ and $\liminf \|\rho^n\|_p \geq \|\rho\|_p$. For the V term we write $m = m_1 + m_2$ with $m_1 = m \chi_R$ and choose R large enough so that $|m_2| < \delta$ (since $|m| = Z < \infty$). If $V_2 = m_2 * |x|^{-1}$ then $\int V_2 |\rho - \rho^n| < \delta(\text{const})$ by Lemma 7.3 (with $\varepsilon = 1$). Next, write $V_1 = V + V_-$ with $V_- = V_1 \chi_{r^c}$. If $r > 2R$, $V_-(x) < 2Z/|x|$. Let Q_r be the uniform charge distribution inside B_r so that $Q_r * |x|^{-1} = 2Z/|x|$ outside B_r . Then $\int V_- |\rho - \rho^n| \leq D(Q_r, Q_r)^{1/2} D(|\rho - \rho^n|, |\rho - \rho^n|)^{1/2}$. Choose r large enough so that $D(Q_r, Q_r) < \delta^2$. $V_- \in L^{3/2}$, so $\int V_- (\rho - \rho^n) \rightarrow 0$. Since δ was arbitrary, $\int V (\rho - \rho^n) \rightarrow 0$. Combining all this, $\liminf \mathcal{G}'(\psi^n) \geq \mathcal{G}'(\psi)$. Finally, if $\int \rho^n \leq \lambda$ then $\int \rho \leq \lambda$ as in the proof of Theorem 2.4 (but using L^3). As remarked in the proof of Theorem 7.2, we can choose $\psi^n(x) \geq 0$ everywhere; hence $\psi(x) \geq 0$ and $\rho(x) = \psi(x)^2$ minimizes $\mathcal{G}(\rho)$. The uniqueness of ρ follows from the strict convexity of $\mathcal{G}(\rho)$. ■

Definition. λ_c can be defined as in Sec. II, namely, $\lambda_c = \sup\{\lambda | E(\lambda) = \lim_{\lambda \rightarrow \infty} E(\lambda)\}$. A simple variational calculation, which exploits the fact that $V(x) \approx -Z/|x|$ for large $|x|$, shows that $\lambda_c > 0$.

Theorem 7.7. *There is a minimizing ρ on F'_λ with $\int \rho = \lambda$ if and only if $\lambda \leq \lambda_c$. The minimizing ρ in Theorem 7.6 when $\lambda > \lambda_c$ is the ρ for λ_c . $E(\lambda)$ is strictly convex on $[0, \lambda]$.*

Proof. Same as for Theorem 2.5. ■

Theorem 7.8. (i) *Any minimizing $\psi \in F'_\rho$ for $\mathcal{G}'(\psi)$ on the set $\int \psi^2 \leq \lambda$ satisfies the TFW equation (in the sense of distributions):*

$$[-\Delta + W_\rho(x)] \psi(x) = -\mu \psi(x), \quad (7.9)$$

where

$$W_\rho(x) = \gamma \psi(x)^{2p-2} - \phi_\rho(x) \quad (7.10)$$

and

$$\phi_\rho = V - |x|^{-1} * \rho \quad \text{with } \rho = \psi^2.$$

(ii) *If ψ minimizes $\mathcal{G}'(\psi)$ on G'_ρ , then ψ satisfies Eq. (7.9) with $\mu = 0$.*

(iii) *$E(\lambda)$ is continuously differentiable and $-\mu = dE/d\lambda$ for $\lambda \leq \lambda_c$, while $0 = dE/d\lambda$ for $\lambda \geq \lambda_c$. In particular, $\mu \geq 0$.*

(iv) *If $\rho \in G_\rho$ satisfies Eq. (7.9) and $\int \rho = \lambda$ (possibly ∞), then ρ minimizes $\mathcal{G}(\cdot)$ on the set $\int \rho \leq \lambda$.*

(v) *Fix λ . There can be at most one pair ρ, μ [with $\rho(x) \geq 0$] that satisfies Eq. (7.9) with $\int \rho = \lambda$.*

Proof. (i) and (ii) are standard. Just consider $\psi + \varepsilon f$ with $f \in C_0^\infty$ and $(f, \psi) = 0$ and set $d\mathcal{G}/d\varepsilon = 0$. For the absolute minimum we do not require $(f, \psi) = 0$. The proof of (iii) is as in Theorem 2.7 (cf. LS Theorem II.10 and Lemma II.27). The proof of (iv) and (v) imitates that of Theorem 2.6. ■

We shall eventually prove that the minimizing ψ is

unique (we already know that ρ , and hence ψ^2 is unique), but Theorem 7.9 is needed first. We also want to prove that $\lambda_c < \infty$, i.e., the ψ that satisfies Eq. (7.9) with $\mu = 0$ satisfies $\int \psi^2 < \infty$. This will be done in Sec. VII.B.

Theorem 7.9. *If $\psi \in G'_\rho$ satisfies Eq. (7.9) (as a distribution) and $\psi(x) \geq 0$ for all x , then (i) ψ is continuous. More precisely, $\psi \in C^{0,\alpha}$ for every $\alpha < 1$ [i.e., for every bounded ball B , $|\psi(x) - \psi(y)| < M|x - y|^\alpha$ for some M and all $x, y \in B$]. (ii) If V is C^∞ on some open set Ω , then ψ is C^∞ on Ω . For point nuclei, $\Omega = \mathbb{R}^3 \setminus \{R_i\}$. (iii) Either $\psi = 0$ or $\psi(x) > 0$ everywhere. (iv) $W_\rho \in L_{\text{loc}}^{3-\varepsilon}$ for every $\varepsilon > 0$.*

Proof. Clearly $V \in L_{\text{loc}}^{3-\varepsilon}$ (all $\varepsilon > 0$), and since $\psi \in L^6$, $-\Delta \psi \leq f$ with $f = V \psi \in L_{\text{loc}}^{2-\varepsilon}$ (all $\varepsilon > 0$). Choosing $\varepsilon < \frac{1}{2}$, we can apply a result of Stampacchia (1965, Theorem 5.2) to conclude $\psi \in L_{\text{loc}}^\infty$ and hence $\psi^{2p-1} \in L_{\text{loc}}^{3-\varepsilon}$ (all $\varepsilon > 0$). Now, $g = |x|^{-1} * \rho \in L^6$ [since $\Delta g = -4\pi\rho \Rightarrow K\|g\|_6^2 \leq \int (\nabla g)^2 = 8\pi D(\rho, \rho)$]. Therefore $-\Delta \psi \in L_{\text{loc}}^{3-\varepsilon}$ (all $\varepsilon > 0$). Then (Adams, 1975, p. 98) $\psi \in C^{0,\alpha}$. (ii) follows by a bootstrap argument as in Theorem 2.8. For (iii) we note that $-\Delta \psi = b\psi$ and $b \in L_{\text{loc}}^q$, $q > \frac{3}{2}$. The conclusion follows from Harnack's inequality (Gilberg and Trudinger, 1977). ■

We know that $\rho^{1/2} \geq 0$ satisfies Eq. (7.9), so $\rho^{1/2}$ enjoys the above properties. Since ρ is unique we shall henceforth denote Eq. (7.10) simply by W . We shall also use the notation

$$H = -\Delta + W. \quad (7.11)$$

Theorem 7.10. *The minimizing ψ is unique up to a sign which is fixed by $\psi(x) = \rho(x)^{1/2} > 0$ everywhere. ψ is also the unique ground-state eigenfunction of $H = -\Delta + W(x)$ and μ is its ground-state eigenvalue.*

Proof. If ψ is minimizing then $\psi^2 = \rho$ and H are uniquely determined. $f = \rho^{1/2}$ satisfies $Hf = -\mu f$. Since f is non-negative, it is the ground state of H , and the ground state of H is unique up to sign (cf. Reed and Simon, 1978, Sec. XIII.12). ■

Remarks. (i) It is not claimed that the TFW equation (7.9) and (7.10) has no solution other than the positive one. Infinitely many other solutions probably exist. They have been found for certain nonlinear equations which have some resemblance to the TFW equation (Berestycki and Lions, 1980), but the TFW equation itself has not been analyzed in this regard. These other solutions correspond, in some vague sense, to "excited states."

(ii) The interplay between $\mathcal{G}'(\psi)$ and $\mathcal{G}(\rho)$ should be noted. Apart from the somewhat pedantic question of the uniqueness of ψ , $\mathcal{G}$ was used to get the uniqueness of $\rho = \psi^2$ and the convexity of $E(\lambda)$. $\mathcal{G}'$ was used to get the TFW equation in which it is not necessary to distinguish between $\rho(x) > 0$ and $\rho(x) = 0$ as in the TF equation (2.18). The ψ of interest automatically turns out to be positive. For purposes of comparison, the TF equation is $(W + \mu)\psi = 0$ if $\psi > 0$, and $(W + \mu) \geq 0$ if $\psi = 0$. The TFW equation is $(W + \mu)\psi = \Delta \psi$ everywhere.

(iii) Note that there is a solution even for $\mu = 0$. For this ρ , $H = -\Delta + W$ has zero as its ground-state eigenvalue with an L^2 eigenfunction, ψ (Theorem 7.12). This is unusual. Zero is also the bottom of the essential spectrum of H .

To complete the picture of $E(\lambda)$ we have to know how $E(\lambda)$ behaves for small λ . Since μ is a decreasing function of λ (by convexity of E), μ has its maximum at $\lambda=0$.

Theorem 7.11. $\mu(\lambda=0) = -e_0$ where $e_0 < 0$ is the ground-state energy of the Hamiltonian $H_0 = -\Delta - V(x)$. In particular, for a point nucleus $\mu(\lambda=0) = z^2/(4A)$.

Proof. $\mu(0) = \lim_{\lambda \rightarrow 0} E(0) - E(\lambda)$. $E(0) = U$. Let f be the normalized ground state of H_0 : $H_0 f = e_0 f$. Let $\bar{\rho} = \lambda f^2$. Clearly, $\mathcal{E}(\bar{\rho}) = \lambda e_0 + U + o(\lambda)$, since $p > 1$. On the other hand, for any ρ with $\int \rho = \lambda$, $\mathcal{E}(\rho) \geq \lambda e_0 + U$. ■

In Sec. VII.B we shall see that $Z < \lambda_c < \infty$. Therefore the behavior of $E(\lambda)$ can be summarized as follows:

$e(\lambda) = E(\lambda) - U$ in TFW theory looks like Fig. 1 with two important changes: (i) $\lambda_c > Z$ (at least for $p \geq \frac{5}{3}$). $e(\lambda)$ is strictly convex for $0 \leq \lambda \leq \lambda_c$. (ii) The slope at $\lambda=0$ is finite. [In TF theory $e(\lambda) \approx \lambda^{1/3}$.]

B. Properties of the density and λ_c

Our main concern here will be to estimate λ_c . For energetic reasons, it is intuitively clear that $\lambda_c \geq Z$ for large enough p because otherwise the energy could always be lowered by adding some additional charge far out. Benguria (1979) proved this for $p \geq \frac{4}{3}$. We shall also see that $\lambda_c > Z$ for $p \geq \frac{5}{3}$.

What is far from obvious, however, is that λ_c is finite. There is no energetic reason why $E(\lambda)$ could not steadily decrease (and be bounded, of course). It is easy to construct a $\rho(x)$ with $\int \rho = \infty$ so that all the terms in the energy and also $\phi(x)$, except at the nuclei, are finite. $\rho(x) = (1+x^2)^{-3/2}$ is an example. That $\lambda_c < \infty$ is a subtle fact. The same question arises in quantum theory, and it has only recently been proved there that λ_c is finite. Ruskai (1981) proved this when the "electrons" are bosons. I. M. Sigal later found a proof (by a different method) for fermions (paper in preparation).

In the following, ψ always means $+p^{1/2}$. For simplicity we shall henceforth assume the following condition in addition to $V \in \mathcal{D}$:

$$V(x) \leq C/|x| \quad (7.12)$$

for some $C < \infty$ and for all $|x| > \text{some } R$. The fact that $V = |x|^{-1} * m$ and $|m| = Z$ does not guarantee Eq. (7.12). If, however, m has compact support, then (7.12) holds.

Theorem 7.12. $\lambda_c < \infty$ for all $p > 1$.

Proof. Let ρ give the absolute minimum of $\mathcal{E}(\rho)$ on G_p . ψ satisfies Eq. (7.9) with $\mu=0$. We shall prove that this ρ has $\lambda = \int \rho < \infty$, thereby proving that $E(\lambda)$ has an absolute minimum at λ , and hence that $\lambda_c = \lambda$. Assume $\lambda = \infty$. Then for $|x| \geq \text{some } R$ [which is bigger than the R in Eq. (7.12)], $|x|^{-1} * \rho > 2C/|x|$. Thus, for $|x| > R$, $-A\Delta\psi \leq -C\psi/|x|$. Now we use a comparison argument. Let

$$f(x) = M \exp\{-2[C|x|/A]^{1/2}\}$$

with $M > 0$. f satisfies $-A\Delta f \geq -Cf/|x|$, for $|x| \neq 0$, so $-A\Delta(\psi - f) \leq -C(\psi - f)/|x|$. Fix M by $f(x) \geq \psi(x)$ for $|x| = R$. If we knew that $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$ we could conclude, from the maximum principle, that $\psi < f$ for $|x| \geq R$. This implies that $\psi \in L^2$. Unfortunately, we only know that $\psi(x) \rightarrow 0$ in a weak sense (namely, L^6). This, it

turns out, is good enough. See BBL for details. ■

Now that we know $\int \rho < \infty$, even for the absolute minimum ($\mu=0$), we can prove

Theorem 7.13. ψ is bounded on $\mathbb{R}^3$ and $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$. Also, $\psi \in H^2$ (i.e., $\psi, \nabla\psi$ and $\Delta\psi \in L^2$).

Proof. $-A\Delta\psi < V\psi$ so $(-A\Delta + 1)\psi \leq (V+1)\psi$. Since $(V+1)\psi \in L^2$, $\psi \leq (-A\Delta + 1)^{-1}[(V+1)\psi]$ and this is bounded and goes to zero as $|x| \rightarrow \infty$ (Lemma 3.1). Finally, $\psi^{2p-1} \leq d\psi$ for some d , and $g = |x|^{-1} * \rho \in L^6$ together with $\psi \in L^6$ imply $g\psi \in L^2$. Hence $\Delta\psi \in L^2$. ■

Theorem 7.14. If $p \geq \frac{5}{3}$ then, for all x ,

$$\gamma\rho(x)^{p-1} \leq V(x). \quad (7.13)$$

In particular, if $p = \frac{5}{3}$, $\rho(x) < [V(x)/\gamma]^{3/2}$.

Proof. The essential point is that since V is superharmonic, so is V' for $t \leq 1$. Let $f = \psi - (V/\gamma)^t$ with $t = 1/(2p-2)$. Let $B = \{x | f(x) > 0\}$. Since ψ and V are continuous on B , B is open. On B , $W > 0$ so $-\Delta f < 0$. $f=0$ at ∞ and on ∂B , so B is empty. ■

Remark. The bound in Eq. (7.13) also holds trivially in TF theory from Eq. (2.18).

Theorem 7.15. If $p \geq \frac{4}{3}$ then $\lambda_c \geq Z$.

Proof. Suppose $\lambda_c = Z - \epsilon$. Since $H = -A\Delta + W$ has zero as its ground-state energy, $(f, Hf) \geq 0$ for any $f \in C_0^\infty$. Let $f_t(x) \neq 0$ be spherically symmetric with support in $1 \leq |x| \leq 2$, $f_t(x) \leq 1$, and $f_n(x) = f(x/n)$. Then $\int f_t^2 \phi = \int f_t^2 [\phi]$, where $[\phi]$ is the spherical average of ϕ . It is easy to see that for $|x| \geq \text{some } R$, $[\phi] \geq \epsilon/2|x|$ since $\int \rho = Z - \epsilon$. Therefore $\int f_t^2 \phi \geq (\text{const})n^2$ for large n . $\int (\nabla f_n)^2 = (\text{const})n$. The crucial quantity is $D_n = \int f_n^2 \rho^{p-1}$. If $p \geq 2$, $D_n \leq (\text{const}) \int \rho$. If $p < 2$ use Hölder's inequality: $D_n \leq X_n^{p-1} Y_n^{2-p}$, where $X_n = \int f_n^2 \rho$ and $Y_n = \int f_n^2$. Clearly $X_n \rightarrow 0$ as $n \rightarrow \infty$ since $\rho \in L^1$. $Y_n = (\text{const})n^3$. Now let $n \rightarrow \infty$, whence $(f_n, Hf_n) \rightarrow -\infty$. ■

Remarks. (i) The basic reason that $p \geq \frac{4}{3}$ is needed in the proof of Theorem 7.15 is that we want to be able to ignore the ρ^{p-1} term in W and thereby obtain a negative-energy bound state for H when $\lambda < Z$. However, if $\rho \in L^1$ then (essentially) $\rho(x) \sim |x|^{-3} f(x)$, where $f(x)$ can be slowly decreasing. Hence we can be certain that ρ^{p-1} is small compared to $|x|^{-1}$ only if $3(p-1) \geq 1$.

(ii) In Theorems 7.16 and 7.19 we prove that $\lambda_c > Z$. The underlying idea is that to have a zero-energy L^2 bound state, $W(x)$ has to be positive for large $|x|$. Essentially, $W(x)$ has to be as big as $|x|^{-2}$; this requirement is clear if we assume $\psi(x) \sim |x|^{-a}$ for large $|x|$. If $\lambda_c = Z$, then ϕ is (essentially) positive for large $|x|$, so the repulsion has to come from ρ^{p-1} . But if $p-1 \geq \frac{2}{3}$ then ρ^{p-1} cannot be sufficiently big since $\rho \in L^1$. The theorem that $\lambda_c > Z$ when $p \geq \frac{5}{3}$ was proved for an atom in BBL. We give that proof first in Theorem 7.16 in order to clarify the ideas. Then, after Lemma 7.18, we give a proof (which is not in BBL) of the general case in Theorem 7.19. Some condition on p really is needed to have $\lambda_c > Z$. In BBL it is proved that if $p = \frac{3}{2}$, $\gamma=1$, V is given by Eq. (2.1), and $A \leq 1/16\pi$, then $\lambda_c = Z$.

Theorem 7.16. Suppose $p \geq \frac{5}{3}$ and suppose $V = |x|^{-1} * m$

where m is a non-negative measure that satisfies the following conditions: (i) m is spherically symmetric; (ii) the support of m is contained in some ball, $B_R = \{x \mid |x| < R\}$. Then $\lambda_c > Z$.

Proof. Assume that $\lambda_c \leq Z$. By Newton's theorem, $\phi(x) \geq 0$ for $|x| > R$. Then when $\lambda = \lambda_c$, $-\Delta\psi \geq -\gamma\psi^{2p-1}$ for $|x| > R$. ψ is spherically symmetric and $\psi(R) > 0$. Let $f(x) = C|x|^{-3/2}$, which satisfies $-\Delta\psi \geq -\gamma\psi^{2p-1}$ for $|x| \geq R$ provided $0 < C \leq D$ with $D^{2p-2} = (3A/4\gamma)R^{3p-5}$. Let $C = \min[D, \psi(R)]$. Then $\psi(x) \geq f(x)$ for all $|x| > R$, because $-\Delta(\psi - f) \geq -\gamma(\psi^{2p-1} - f^{2p-1})$, which would imply that $\psi - f$ is superharmonic on the set where $\psi - f < 0$. Since ψ and f go to zero at infinity, and $\psi - f \geq 0$ at $|x| = R$, this is impossible. Hence $\psi \notin L^2$, which contradicts $\|\psi\|_2 \leq Z$. ■

In the foregoing we used a comparison argument which, in turn, relied on the fact that the positive part of W , namely ρ^{p-1} , was simply related to ψ . In the proof of Theorem 7.19 we shall not have that luxury, and so the more powerful Lemma 7.18 is needed.

Lemma 7.17. Let S_R denote the sphere $\{x \mid |x| = R\}$ and let $d\Omega$ be the normalized, invariant, spherical measure on S_1 . For any function h , let $[h](r) = \int h(r, \Omega) d\Omega$ be the spherical average of h . Now suppose $\psi(x) > 0$ is C^2 in a neighborhood of S_R . Let $f(r) = \exp\{\ln\psi(r)\}$. Then, for all r in some neighborhood of R ,

$$[\Delta\psi/\psi](r) \geq (\Delta f/f)(r) \\ = \{d^2f/dr^2 + (2/r)df/dr\}/f(r).$$

Proof. Let $g(x) = \ln\psi(x)$. Then $\Delta\psi/\psi = \Delta g + (\nabla g)^2$. Clearly $[\Delta g] = \Delta[g]$. Moreover, $(\nabla g)^2 \geq \partial g(r, \Omega)/\partial r^2$, and $[(\partial g/\partial r)^2] \geq (d[g]/dr)^2$ by the Schwarz inequality. Thus $[\Delta\psi/\psi] \geq [\Delta g] + (\nabla[g])^2 = \Delta f/f$. ■

Lemma 7.18. Suppose $\psi(x) > 0$ is a C^2 function in a neighborhood of the domain $D = \{x \mid |x| > R\}$ and ψ satisfies $[-\Delta\psi + W(x)]\psi(x) \geq 0$ on D . Let $[W]$ be the spherical average of W and write $[W] = [W]_+$ with $[W]_-(x) = \max\{[W](x), 0\}$. Suppose $[W]_+ \in L^{3/2}(D)$. Then $\psi \notin L^2(D)$. (Note: no hypothesis is made about $[W]_-$.) See note added in proof below.

Remarks. Simon (1981, Appendix 3) proves a similar theorem for $D = \mathbb{R}^3$, except that $[W] = [W]_+$. $[W]_-$ is replaced by $W = W_+ - W_-$ with $W_+ = \max(W, 0)$. Simon does not require the technical restrictions that $\psi(x) > 0$ and ψ is C^2 . Simon's theorem will be used in our proof. Lemma 7.18 improves Simon's result in two ways: (i) It is sufficient to consider D and not all of $\mathbb{R}^3$. (ii) It is only necessary that $[W]_+$, and not W_+ , be in $L^{3/2}$, the latter distinction is important. As an example, suppose that for large $|x|$ the potential W is that of a dipole, i.e., $W(x_1, x_2, x_3) = x_1|x|^{-3}$. $W_+ \notin L^{3/2}$ but, since $[W]_+ = 0$, Lemma 7.18 says that this W cannot have a zero-energy L^2 bound state.

Proof. Let $f = \exp\{\ln\psi\}$ as in Lemma 7.17. Then $-\Delta f/f + [W] \geq [-\Delta\psi/\psi + W] \geq 0$. By Jensen's inequality $\int f^2 \leq \int \psi^2$, so if $f \notin L^2$ then $\psi \notin L^2$. Therefore it suffices to consider $\{-\Delta f + [W](x)\}f \geq 0$ and to prove $f \notin L^2$ under the stated condition on $[W]$. First, suppose $D = \mathbb{R}^3$. Then this is just Simon's (1981) theorem. (How-

ever, since we are now dealing with spherically symmetric $[W]$ and f , it is likely that a direct, ordinary differential equation proof can be found to replace Simon's proof.) Next, suppose $R > 0$. Let $g(x) > 0$ be any C^2 function defined in $\mathbb{R}^3$ such that $g(x) = f(x)$ for $|x| \geq R$. Then $\{-\Delta g + U(x)\}g \geq 0$ on $\mathbb{R}^3$ where $U = [W]$ for $|x| \geq R$ and U is bounded for $|x| \leq R$. Clearly $[W]_+ \in L^{3/2}(D)$ if and only if $U \in L^{3/2}(\mathbb{R}^3)$. Apply Simon's theorem to U . ■

Note added in proof. H. Brezis (private communication) has found a direct ordinary differential equation proof. Moreover, under the hypotheses of Lemma 7.18, $\psi \notin L^{3-\epsilon}$ for all $\epsilon > 0$.

Theorem 7.19. Let the hypothesis be the same as in Theorem 7.16 except that (i) is omitted. (In other words, a molecule is now being considered.) Then $\lambda_c > Z$.

Proof. For $|x| > R$, $V(x)$ is C^∞ so $\psi(x) > 0$ and $\psi \in C^2$ by Theorem 7.9. Assume $\lambda_c \leq Z$. The hypotheses of Lemma 7.18 are satisfied with $[-\Delta\psi + W(x)]\psi = 0$. To obtain a contradiction we have to show $[W]_+ \in L^{3/2}$. Consider ϕ . Even if ϕ is negative somewhere, $[\phi](r) > 0$ in D by Newton's theorem. Therefore it suffices to show $[\rho^{p-1}] \in L^{3/2}$. If $p \geq \frac{3}{2}$ then $p-1 \geq \frac{1}{2}$ and $[\rho^{p-1}](r) < C[\rho^{2/3}](r)$, since ρ is bounded. But $[\rho^{2/3}](r) \leq \{[\rho](r)\}^{2/3}$, by Hölder, and $\int [\rho]^{(2/3)(3/2)} = \int \rho < \infty$. ■

We know that $\lambda_c > Z$ in the physically interesting case $p = \frac{3}{2}$. How large is $\lambda_c - Z$? In other words, how negative can ions be? This seems to be a very difficult question, even for an atom. To obtain qualitative agreement with quantum theory, it would be desirable if $\lambda_c - Z \sim 1$, at least for Z up to 100, say. The only available bound, at present, is Theorem 7.23. First Lemmas 7.20, 7.21, and 7.22 are needed. The lemmas were inspired by the work of R. Benguria (private communication), who proved the lemmas and Theorem 7.23 in the spherically symmetric case (which corresponds to the atom in TFW theory).

Lemma 7.20. Let ψ and f be two real valued functions on $\mathbb{R}^3$ which satisfy $-\Delta\psi = f$ in the sense of distributions. Let r denote the function $|x|$. Suppose $\psi \in L^2$, $f \in L^2$, and $r\psi f \in L^1$. Then, for any constant $d \geq 0$,

$$\int \langle |x|^2 + d \rangle^{1/2} \psi(x) f(x) dx \geq 0.$$

Proof. Using dominated convergence, it is sufficient to consider only $d > 0$. Let $R = (r^2 + d)^{1/2} \in C^\infty$. We have $\Delta R = 2R^{-1} + dR^{-3}$ and $|(R/r)\nabla R|^2 = 1$. Suppose $\phi \in C_0^\infty$ (infinitely differentiable functions of compact support). We claim $I = -\int R\phi\Delta\phi \geq 0$. To see this, integrate by parts: $I = A + B$ with $A = \int (\nabla\phi)^2 R$ and

$$B = \int \phi \nabla\phi \cdot \nabla R = \int (\nabla\phi \cdot \nabla R)(R/r)(r/R)\phi.$$

By Schwarz, and $\{(\nabla\phi \cdot \nabla R)(R/r)\}^2 \leq (\nabla\phi)^2$, we have $B^2 \leq AC$ with

$$2C = 2 \int \phi^2 r^2 R^{-3} \leq \int \phi^2 \Delta R.$$

However,

$$2B = \int \nabla\phi^2 \cdot \nabla R = - \int \phi^2 \Delta R,$$

and hence $|B| \leq A$, which proves the lemma. Now, suppose ψ and $f \notin C_0^\infty$ have compact support. Given $\varepsilon > 0$ there exists $g \in C_0^\infty$ such that $\|\psi - g\|_2 < \varepsilon$, $\|\nabla\psi - \nabla g\|_2 < \varepsilon$, and $\|\Delta\psi - \Delta g\|_2 < \varepsilon$. (Note: since ψ and $\Delta\psi \in L^2$, so is $\nabla\psi$.) Then $gR \in C_0^\infty$ and

$$\begin{aligned} \int gRf &= - \int gR\Delta\psi - \int \psi\Delta(Rg) \\ &= - \int g\Delta(Rg) - M = - \int Rg\Delta g - M, \end{aligned}$$

with $M = \int (\psi - g)\Delta(Rg)$. It suffices to show that $M \rightarrow 0$ as $\varepsilon \rightarrow 0$ because $\int gRf \rightarrow \int \psi Rf$. But

$$M = \int (\psi - g) \{ g\Delta R + 2\nabla g \cdot \nabla R + R\Delta g \}.$$

We can assume $\text{supp}(g)$ is in some fixed ball, independent of ε . Since g , ∇g , and Δg are uniformly L^2 bounded, $M \rightarrow 0$. For the general case, let $h \in C_0^\infty$ satisfy $1 \geq h \geq 0$, $(\nabla h)(0) = 0$, $h(0) = 1$, $h(x) = 0$ for $|x| > 1$. Let $h_n(x) = h(x/n)$ and $\psi_n = h_n\psi$. Then, as a distribution,

$$-\Delta\psi_n = h_n f - 2\nabla h_n \cdot \nabla\psi - \psi\Delta h_n \equiv K_n.$$

By the previous result, $T_n \equiv \int R\psi_n K_n \geq 0$. But $\int h_n^2 \psi f R \rightarrow \int \psi f R$ by dominated convergence. $Rh_n \Delta h_n = n^{-1}P_n(x/n)$ with

$$P_n(x) = (|x|^2 + dn^{-2})^{1/2} (h\Delta h)(x) < a,$$

so $\int R\psi_n \Delta h_n \rightarrow 0$, since $\psi \in L^2$. Similarly, $R\nabla h_n^2 \equiv L_n$ is uniformly bounded and converges pointwise to zero. Since ψ and $\nabla\psi \in L^2$, $\int \psi \nabla\psi L_n \rightarrow 0$ by dominated convergence. ■

Remarks. (i) Lemma 7.20 is useful for L^2 solutions to the Schrödinger equation $[-\Delta + W(x)]\psi = -\mu\psi$. Then $\int \psi^2(W + \mu)r \leq 0$ under some mild conditions on W [e.g., $W(x) < c/r$ for large r , $W \in L_{\text{loc}}^2$ and $r\psi^2 \in L^1$ if $\mu \neq 0$].

(ii) The essential properties of R that were used were $R > 0$ and $R\Delta R \geq 2(\nabla R)^2$. Therefore Lemma 7.20 will hold for functions other than $(r^2 + d)^{1/2}$ having these properties. Formally, this means that if $R(x) = 1/V(x)$ then we require $V > 0$ and $\Delta V < 0$. We state this as Lemma 7.21, whose proof imitates the proof of Lemma 7.20.

Lemma 7.21. Suppose $V = |x|^{-1} * m$, with m a nonnegative measure, $|m| = Z < \infty$, and V satisfies Eq. (7.12). Then if ψ and f satisfy the hypotheses of Lemma 7.20,

$$\int dx \psi(x)f(x)/V(x) \geq 0.$$

Lemma 7.22. Let $\rho(x) \geq 0$, $\int \rho = \lambda < \infty$, and $V = |x|^{-1} * m$ where m is any non-negative measure with $|m| = Z < \infty$. Then

$$I \equiv \int \rho(x)\rho(y)|x-y|^{-1}V(x)^{-1}dx dy \geq \lambda^2/2Z.$$

Proof. Take $Z=1$ and let $0 < \varepsilon < 1$. By Proposition 3.24, $\rho = \rho_1 + \rho_2$ with $\rho_1, \rho_2 \geq 0$, and $H_1 = |x|^{-1} * \rho_1$ satisfies $H_1 \leq \varepsilon\lambda V$ and $H_1 = \varepsilon\lambda V$ when $\rho_2 > 0$. Clearly, $\int \rho_1 \leq \varepsilon\lambda$ by Lemma 3.3. Then $\int \rho_2 \geq (1-\varepsilon)\lambda$ and

$$I \geq \int \rho_2(H_1 + H_2)/V \geq \varepsilon(1-\varepsilon)\lambda^2 + \int \rho_2 H_2/V.$$

Repeat the argument with ρ_2 [using $\int \rho_2 \geq (1-\varepsilon)\lambda$], and

so on *ad infinitum*. Then

$$I \geq \lambda^2 \sum_{j=0}^{\infty} (1-\varepsilon)^{2j} = \lambda^2(1-\varepsilon)/(2-\varepsilon).$$

Now let $\varepsilon \rightarrow 0$. ■

Remark. Benguria proved Lemma 7.22 when $V = 1/r$. In this case one can simply use the fact that $(|x| + |y|)|x-y|^{-1} \geq 1$.

Theorem 7.23. Assume V satisfies Eq. 7.12. Then $\lambda_c < 2Z$, for all $p > 1$.

Proof. We know $\lambda_c < \infty$. Let ψ be the minimizing solution of Eq. (7.9) with $\mu = 0$. Then, by Theorem 7.13, ψ and $f = (\gamma\rho^{p-1} - \phi)\psi$ satisfy the hypotheses of Lemma 7.21. Thus $0 < \int \rho\phi/V = \lambda_c - I$ with $I = \int H\rho/V$ and $H = |x|^{-1} * \rho$. But $I \geq \lambda_c^2/2Z$. ■

Remark. This bound does not involve the value of A in Eq. (7.1). It also does not utilize the $\gamma\rho^{p-1}$ term in W . There is considerable room for improvement.

The next two theorems are about the asymptotics of ψ .

Theorem 7.24. Let ψ be the positive solution to Eq. (7.9), for any p .

(i) Let $\mu > 0$. Then for every $t < \mu$ there exists a constant M such that

$$\psi(x) \leq M \exp[-(t/A)^{1/2}|x|].$$

(ii) Let $\mu = 0$ (i.e., $\lambda = \lambda_c$), and assume $\lambda_c > Z$, as is certainly the case when $p \geq \frac{3}{2}$. Assume also that m has compact support. Then for every $t < \lambda_c - Z$ there is a constant M such that

$$\psi(x) \leq M \exp[-2(t|x|/A)^{1/2}].$$

Proof. (i) is standard. Since ψ and $V \rightarrow 0$ as $|x| \rightarrow \infty$, we have $\psi = -(-\Delta + t)^{-1}(W + \mu - t)\psi$. For $|x| > \text{some } R$, $W + \mu - t > 0$. Therefore, since $\psi > 0$,

$$\psi(x) \leq \int_{|y| \leq R} Y(x-y)[W(y) + \mu - t]\psi(y)dy,$$

where

$$Y(x) = (4\pi A|x|)^{-1} \exp[-(t/A)^{1/2}|x|].$$

The proof of (ii) is the same as the proof of Theorem 7.12. It is only necessary to note that, since m has compact support (in B_R , say), $V(x) \leq Z/(|x| - R)$ for $|x| > R$, and this is $\leq (Z + \varepsilon)/|x|$ for $|x|$ large enough. ■

The next theorem is the well known cusp condition (Kato, 1957).

Theorem 7.25. Let $V(x) = \sum_j |x - R_j|^{-1}$ be the potential of point nuclei. Then at each R_j

$$z_j \psi(R_j) = -2A \lim_{r \rightarrow 0} \int_{|x - R_j| = r} r^{-1}(x - R_j) \cdot \nabla \psi(x) d\Omega,$$

where $d\Omega$ is the normalized uniform measure on the sphere. This holds for any λ . In particular, for an atom with nuclear charge z located at the origin, ψ is spherically symmetric and

$$z\psi(0) = -2A \lim_{r \rightarrow 0} (d\psi/dr)(r).$$

Proof. Recall that, by Theorem 7.9, ψ is C^∞ away from

the R_j and ψ is Hölder continuous everywhere. The theorem is proved by integrating Eq. (7.9) in a small ball B_r and then integrating by parts. The spherical symmetry in the atomic case is implied by uniqueness. ■

Theorem 7.26. Let $V(x) = z/|x|$ be the potential of an atom with a point nucleus. Then, for any λ , $\psi(r)$ is a strictly decreasing function of r .

Proof. In Theorem 7.12 and the remark following it we proved this for $\lambda \leq z$ by using rearrangement inequalities. Here we give a different proof for $\lambda \leq z$ which extends to $\lambda > z$. Recall that ψ is continuous and positive and that ψ is C^∞ for $r > 0$. Also, $\mu > 0$. Let $Q(r) = \int \chi_r \rho$ be the electronic charge inside the ball B_r . By Newton's theorem, the potential ϕ satisfies:

- (i) $\phi(r) \leq [z - Q(r)]/r$.
- (ii) $\dot{\phi} = [Q(r) - z]/r^2$ (dots denote d/dr).
- (iii) If $\lambda \leq z$, $\phi(r) \geq 0$ and $\phi(r)$ is decreasing.
- (iv) If $\lambda > z$ there is a unique $R > 0$ such that $\phi(r) \geq 0$ and decreasing for $r \leq R$, and $\phi(r) < 0$ for $r > R$. $Q(R) < z$.

$\lambda \leq z$: By Theorem 7.24 $\dot{\psi}(r) < 0$ near $r = 0$. If ψ is not monotone, then since $\psi(r) \rightarrow 0$ as $r \rightarrow \infty$, there are two points $0 < r_1 < r_2$ such that $\psi(r_1) \leq \psi(r_2)$, $\dot{\psi}(r_1) \geq 0$, $\dot{\psi}(r_2) \leq 0$, and $\psi(r_1) = \psi(r_2) = 0$. Since ψ does not have compact support, $Q(r) < z$, all r . Hence $W(r_2) > W(r_1)$. Since

$$\dot{\psi}(r_1) = [W(r_1) + \mu] \psi(r_1) \geq 0,$$

we have $W(r_1) + \mu \geq 0$. But then

$$0 \geq \dot{\psi}(r_2) = [W(r_2) + \mu] \psi(r_2)$$

is impossible.

$\lambda > z$: There is an $\varepsilon > 0$ such that $W(r) > 0$ for $r > R - \varepsilon$. Let $D_r = \{x \in \mathbb{R}^3 \mid |x| > r\}$. Take $r > R - \varepsilon$. On D_r , $-\Delta\psi < 0$. Since $\psi > 0$ is subharmonic on D_r , ψ has its unique maximum on ∂D_r , namely, $|x| = r$. This proves the theorem on the domain $\{r \mid r > R - \varepsilon\}$. To prove the theorem on the domain $\{r \mid 0 \leq r < R\}$ the argument in the $\lambda \leq z$ case can be used, since $Q(r) < z$ in this domain. ■

Conjecture. In the point nucleus, atomic case ψ is convex, possibly even log convex.

C. Binding in TFW theory

In TF theory binding never occurs when the repulsion U is included. In TFW theory binding is a common phenomenon. We conjecture that every neutral system (molecule or atom) binds to every other neutral system. In the following, the occurrence of binding will be proved in enough cases to render the conjecture plausible. It will also be seen that binding in TFW theory is intimately connected with the existence of negative ions, i.e., $\lambda_c > Z$. We shall assume here that $\lambda_c > Z$ for all the systems under consideration. $p \geq \frac{2}{3}$ guarantees this, but no requirement on p other than $\lambda_c > Z$ will be made. $V = |x|^{-1} * m$, with m a non-negative measure of compact support [so that Eq. (7.12) is satisfied], $|m| = Z$, and m spherically symmetric in the atomic case.

First, let us define what binding means. Suppose we have two systems (not necessarily atoms) with potentials V_1 and V_2 and a combined system with $V(x) = V_1(x) + V_2(x - R)$ for some vector R . The combined system is

neutral, i.e., $\lambda = Z \equiv Z_1 + Z_2$. Let $E(R)$ denote the energy of the combined system and $E_i(\lambda)$ denote the energies of the subsystems with arbitrary electron charge λ and $E_i = E_i(\lambda = Z_i)$ (note the difference in notation). Then

$$E(\infty) \equiv \min_{0 \leq \lambda \leq Z} E_1(\lambda) + E_2(Z - \lambda). \quad (7.14)$$

Let μ_i be the chemical potentials of the subsystems when they are neutral, i.e., $\lambda_i = Z_i$. We know that $\mu_i > 0$. If $\mu_1 = \mu_2$ then $E(\infty) = E_1 + E_2$. Otherwise, $E(\infty) < E_1 + E_2$. In general, λ in Eq. (7.14) is determined by $\mu_1(\lambda) = \mu_2(Z - \lambda)$ if this equation has a solution for $0 \leq \lambda \leq Z$; otherwise, $\lambda = 0$ if $\mu_1(0) \leq \mu_2(Z)$ and $\lambda = Z$ if $\mu_2(0) \leq \mu_1(Z)$. (Recall that $\mu_i(\lambda)$ is monotone.)

If $\mu_1 \neq \mu_2$ then the subsystems *spontaneously ionize* when they are infinitely far apart. This is *not* considered to be binding. For real atoms the phenomenon of spontaneous ionization apparently never occurs, because it seems to be the case that the lowest ionization potential among all atoms is less than the largest electron affinity. (I thank J. Morgan III for pointing this out to me.) In real atoms, λ and $Z - \lambda$ in Eq. (7.14) are restricted to be integral, but no such restriction occurs in TFW theory. In TF theory the phenomenon never occurs because μ_i is always zero.

In TFW theory it is possible for λ to be zero in Eq. (7.14), i.e., one subsystem is completely stripped of electrons. Let $V_i = z_i/r$ with $z_2 \gg z_1$. If $\mu_1(\lambda = 0) \leq \mu_2(\lambda = z_1 + z_2)$, then $\lambda = 0$ in Eq. (7.14). By Theorem 7.11, $\mu_1(\lambda) \leq \mu_1(0) = z_1^2/4A$. Since $\lambda_c(2) > z_2$ and $\mu_2 > 0$, the above inequality will hold for any fixed z_2 if z_1 is chosen small enough. This case was cited in BBL as an example where binding occurs (see Theorem 7.27).

Definition. Binding is said to occur if $E(R) < E(\infty)$ for some R .

Theorem 7.27. Suppose the chemical potentials of the neutral subsystems are unequal, i.e., $\mu_1 \neq \mu_2$. Then binding occurs. (This holds for all $p > 1$, even if $\lambda_c = Z$ for one or more of the three systems.)

Proof. Suppose $\lambda < Z_1$ in Eq. (7.14). Then when $R = \infty$ subsystem 1 is positively charged (with charge $Q = Z_1 - \lambda$) and subsystem 2 has charge $-Q$. Let ρ_i be the TFW densities (with λ and $Z - \lambda$, respectively). By Theorem 7.24, ρ_i has exponential falloff. For the combined system (at R) consider the variational ρ defined by $\rho(x) = \rho_1(x) + \rho_2(x - R)$. The first term in Eq. (7.1) is sub-additive (by convexity). For large R the total Coulomb energy decreases essentially by $-Q^2/R$ because of the exponential falloff of each ρ_i . The $\int \rho^p$ term is super-additive, but it increases only by a term of order $\exp[-(\text{const})R]$ for large R . We omit the easy proof of these last two assertions. Thus for large enough, but finite, R , $E(R) < E(\infty)$. ■

The difficult case is $\mu_1 = \mu_2$. Henceforth we confine our attention to atoms.

Conjecture. If $z_1 < z_2$ for two atoms with point nuclei, then $\mu_1(\lambda = bz_1) < \mu_2(\lambda = bz_2)$ for all $b \leq 1$. In particular $\mu_1 < \mu_2$. Moreover, $\lambda_c(1) - z_1 < \lambda_c(2) - z_2$.

If this conjecture is correct then only the homopolar case has to be considered for point nuclei. In Theorem 7.28 we prove binding for the homopolar molecule, even

for "smeared" nuclei. However, we have already shown that binding occurs if $z_1 \ll z_2$, so it is likely that binding always occurs, even if the conjecture is wrong.

Theorem 7.28. *Binding occurs for two equal atoms for any nuclear charge z and for any $p > 1$ provided $\lambda_e > z$ for the atom.*

Proof. We shall construct a variational $\tilde{\rho}$ for the combined system, with $\int \tilde{\rho} = 2z$, such that $\mathcal{E}(\tilde{\rho})$ for the combined system at some R is less than $E(\infty) = 2E_1$. First, consider the atom with the nucleus at the origin and with $\lambda = z + \varepsilon$, where $z < \lambda < \lambda_e$. Let ρ be the TFW density. Denote E_1 by E and μ_1 by μ . Center the nucleus at the point $(-R, 0, 0)$, where $R > 0$, depending on ε , is such that $\int \chi_- \rho = z$, with χ_- being the characteristic functions of the half space, $H = \{(x_1, x_2, x_3) | x_1 \leq 0\}$. Assume, for the moment, that the nuclear m has support in $B_{R/2}$, i.e., the displaced m has support in $\{x_1 \leq -R/2\}$. Center the second atom at $(R, 0, 0)$. Its corresponding density is ρ^* , where the asterisk means reflection through the plane $x_1 = 0$. Choose the variational $\tilde{\rho} = \rho_- + \rho_+^*$ with $\rho_- = \chi_- \rho$. Clearly $\tilde{\rho}$ is continuous across the plane $x_1 = 0$, and $\int \tilde{\rho} = 2z$, so it is a valid variational function. In the following bookkeeping of $\mathcal{E}(\tilde{\rho})$ we use the terminology "energy gain" (resp. "loss") to mean that the contribution to $\mathcal{E}(\tilde{\rho})$ is negative (resp. positive) relative to $2E$. Before the χ_- cutoff, we start with $2E_1(\lambda) \leq 2E - (2\varepsilon)(\mu/2)$ if ε is small enough, so we have gained $\varepsilon\mu$. This linear term in ε is the crucial point; it exists because $\lambda_e > z$. After the cutoff we gain the kinetic energy [first two terms in Eq. (7.1)] contributions from the missing pieces of ρ and ρ^* . Next, we lose on the $-\int V\rho$ term (for each atom separately) because of the missing pieces. Each missing charge is ε and its distance to its atomic origin is R . Since the atomic $V(r) \leq z/r$, the energy loss is at most $2(\varepsilon z/R)$. Clearly we gain on the missing atomic repulsion, $D(\rho, \rho)$ term. Finally, if $dM(x) = dm(x + R) - \rho^*(x)dx$ is the total charge density in H , we lose the atom-atom interaction $\Delta = 2D(M, M^*)$. By reflection positivity, $\Delta \geq 0$. (See Benguria and Lieb (1978b), Lemma B.2.) On balance, the net energy gain is at least $\varepsilon(\mu - 2z/R) - \Delta$.

Now we claim two things: (i) As $\varepsilon \rightarrow 0$, $R \rightarrow \infty$. (ii) $\Delta < C\varepsilon z/R$ for some constant C . [Actually, it is possible to prove $\Delta < o(\varepsilon)z/R$.] Using (i) and (ii) we are done, because for sufficiently small ε the gain is positive and the assumption on $\text{supp}(m)$ is justified.

Proof of (i). Let ρ_n be the atomic density for $\lambda = z + \varepsilon_n$, with $\varepsilon_n \rightarrow 0$. As in the proof of Theorem 7.6, we can find a weakly convergent subsequence, $\rho_n \rightharpoonup \tilde{\rho}$, so that $E = \lim \mathcal{E}(\rho_n) \geq \mathcal{E}(\tilde{\rho})$. But $\int \tilde{\rho} \leq z$, so $\tilde{\rho}$ must be ρ_e , the atomic density with $\lambda = z$. If R_n does not tend to ∞ then, for large enough n , $\int \chi_- \rho_n < z$ by the weak convergence, which is a contradiction.

Proof of (ii). This is messy. Let B_r be the ball of radius r centered at $(-R, 0, 0)$ and let χ_r be its characteristic function. Write $\rho = \rho^a + \rho^b$, where $\rho^a = \chi_{3R/2}\rho$. By elementary geometry, $d \int \rho^b \geq \int \rho^b$ with $d < 1$. Since

$$\int \rho^b(1 - \chi_-) < \varepsilon, \quad \int \rho^b < \varepsilon/(1 - d).$$

Let $t = d/(1 - d)$. The contribution of ρ^b to Δ is

$$\begin{aligned} -4D(\rho_-^b, M^*) - 2D(\rho_-^b, \rho_-^{b*}) &\leq -4D(\rho_-^b, M^*) \leq 4D(\rho_-^b, \rho_-^{b*}) \\ &\leq 4D(\rho^b, \rho_-^{b*}) \leq 2z \left(\int \rho^b \right) / R, \end{aligned}$$

since the potential of ρ^b is everywhere less than $(3R/2)^{-1} \int \rho^b$. Henceforth we can assume $\rho_- = \rho_-^a$ and $z > \int \rho_-^a > z - t\varepsilon$. This assumption changes M to M^a . Let $d\tilde{M}(x) = dm(x + R) - \rho^a(x)dx$. [Note: $\text{supp}(\tilde{M})$ extends outside H , but is inside $\{x_1 \leq R/2\}$.] $\phi = |x|^{-1} * M^a$ is subharmonic on $\text{supp}(\tilde{M})$ and harmonic on $\text{supp}(m)$ so

$$D(\tilde{M}, M^a) \leq \left(\int d\tilde{M} \right) D(\delta, M^a),$$

where δ is a delta function at $(-R, 0, 0)$. This is

$$\leq \left| \int d\tilde{M} \right| z/R \leq t\varepsilon z/R,$$

since the distance of $\text{supp}(M^a)$ to $(-R, 0, 0)$ is R . Finally,

$$\begin{aligned} D(M^a - \tilde{M}, M^a) &= D(\rho^a - \rho_-^a, M^a) \leq D(\rho^a - \rho_-^a, m^*) \\ &= D(\rho^a - \rho_-^a, z\delta^*) \leq 2\varepsilon z/R \end{aligned}$$

since $\int \rho^a - \rho_-^a \leq \varepsilon$ and the distance of $\text{supp}(\rho^a)$ to $(R, 0, 0)$ is $R/2$. ■

I thank J. Morgan III for valuable discussions about Theorem 7.28. Balázš (1967) gave a heuristic argument for the binding of two equal atoms with point nuclear charges.

D. The Z^2 correction and the behavior near the nuclei

Here we consider point nuclei with potential given by Eq. (2.1). The question we address is what is the principal correction to the TF energy and density caused by the first term in Eq. (7.1)? This term, $A \int (\nabla \rho^{1/2})^2$, will henceforth be denoted by T . For simplicity we confine our attention to $\rho = \frac{5}{3}$, the physical value of ρ .

$E^{\text{TF}} \sim T^{2/3}$. In particular, for a neutral atom,

$$E^{\text{TF}} = -3.67874z^{7/3}/\gamma \quad (7.15)$$

(I thank D. Liberman for this numerical value). At first sight, it might be thought that the leading energy correction is $O(z^{5/3})$. If $\rho^{\text{TF}}(z, r) = z^{2/3} \rho^{\text{TF}}(1, z^{1/3}r)$ is inserted into T , then, by scaling, $T(z) = z^{5/3} T(z=1)$. But $T(z=1) = \infty$ since $\rho^{\text{TF}} \sim r^{-3/2}$ for small r . Thus, for point nuclei, T cannot be regarded as a small perturbation.

The actual correction is $+O(z^2)$ and bounds of this form can easily be found. The following bounds are for an atom, and can obviously be generalized for molecules.

Upper bound: Use a variational $\tilde{\rho}_{\text{TF}}$ for TFW of the form $\tilde{\rho}(r) = \rho^{11}(r)$ for $r > 1/z$ and $\tilde{\rho}(r) = \rho^{11}(1/z)$ for $r < 1/z$.

Lower bound: Let $b > 0$ and write $V(r) = \tilde{V}(r) + H(r)$, where $H(r) = z/r - z^2/b$ for $zr < b$ and $H(r) = 0$ otherwise. For small enough b , $-\Delta\Delta + H > 0$, since $\|H\|_{3/2} \sim b$. Now $\tilde{V} = |x|^{-1} * m$, with $m \geq 0$ and $|m| = z$. Let $\tilde{\rho}$ minimize $\mathcal{E}^{11}(\tilde{V}, \rho)$ with energy $E^{11}(\tilde{V})$. Then $E^{11;W} \geq E^{11}(\tilde{V})$. But $E^{11}(V) \leq \mathcal{E}^{11}(V, \tilde{\rho}) = E^{11}(\tilde{V}) - \int \tilde{\rho}H$. It is not hard to prove, from the TF equation with $\tilde{V}$, that this last integral is $O(z^2)$.

The foregoing calculations show that the main correc-

tion in TFW theory comes from distances of order z^{-1} near the nuclei. The calculations, if carried out for arbitrary λ , also show that the correction is essentially independent of λ . We now show how this correction can be exactly computed to leading order in z , namely, $O(z^2)$.

Let us begin by considering the atom without electron-electron repulsion. The TF theory of such an atom was presented in Sec. V.B following Eq. (5.30). The analogous TFW equation (with $\delta = A\hbar^2/2m$ and $\hbar^2/2m = 1$) is

$$[-\Delta + W(x)]\psi = -\mu\psi, \quad (7.16)$$

with $W(x) = \gamma\rho(x)^{2/3} - z/|x|$, and $\rho = \psi^2$. The absolute minimum, which corresponds to $\lambda = \infty$, has $\mu = 0$, namely,

$$(-\Delta + \gamma|\psi|^{4/3} - z|x|^{-1})\psi = 0. \quad (7.17)$$

The first task is to analyze Eq. (7.17). By simple scaling, any solution scales with A , γ , and z as

$$\psi(z, \gamma, A; x) = (z^2/A\gamma)^{3/4}\psi(1, 1, 1; zx/A). \quad (7.18)$$

Up to Eq. (7.28) we take $z = \gamma = A = 1$. Consider the functional

$$\mathcal{F}'(\psi) = T(\psi) + P(\psi), \quad (7.19)$$

$$T(\psi) = \int (\nabla\psi)^2, \quad P(\psi) = \int k(\psi(x), x)dx, \quad (7.20)$$

$$k(\psi, x) = 3|\psi|^{10/3}/5 + 2|x|^{-5/2}/5 - |\psi|^2|x|^{-1}. \quad (7.21)$$

Note that $k \geq 0$ and, for each x , k has a minimum at $\psi = |x|^{-3/4}$. The function space for $\mathcal{F}'$ is

$$G' = \{\psi | \nabla\psi \in L^2, P(\psi) < \infty\}. \quad (7.22)$$

G' is not convex since $0 \notin G'$. Clearly, Eq. (7.17) is the variational equation for $\mathcal{F}'$. We can also define $G = \{\rho | \rho \geq 0, \rho^{1/2} \in G'\}$ and $\mathcal{F}(\rho) = \mathcal{F}'(\rho^{1/2})$. G is convex and $\rho \rightarrow \mathcal{F}(\rho)$ is convex.

Theorem 7.29. $\mathcal{F}'(\psi)$ has a minimum on G' . This minimizing ψ is unique, except for sign, and satisfies: (i) $\psi > 0$. (ii) ψ is spherically symmetric. (iii) ψ satisfies Eq. (7.17). (iv) ψ is the only non-negative solution to Eq. (7.17) in G' . (v) ψ is C^∞ for $|x| > 0$. (vi) ψ satisfies the cusp condition $2(d\psi/dr)(0) = -\psi(0)$. (vii) for large $r = |x|$, ψ has the asymptotic expansion [which can be formally deduced from Eq. (7.17)],

$$\psi(r) = r^{-3/4} - \frac{9}{64}r^{-7/4} - \frac{3}{64}(21)2r^{-11/4} + O(r^{-15/4}), \quad (7.23)$$

$$\rho(r) = r^{-3/2} - \frac{9}{32}r^{-5/2} - (621/2^{11})r^{-7/2} + O(r^{-9/2}). \quad (7.24)$$

(viii) Any solution f to Eq. (7.17) in G' satisfies $|f(x)| < |x|^{-3/4}$. (ix) By (viii), ψ is superharmonic, and thus $\psi(r)$ is decreasing.

The proof of Theorem 7.29 follows the methods of Secs. VII.A and VII.B, and is given in Lieb, 1981b. The following numerical values, together with a tabulation of ψ , are in Liberman and Lieb, 1981. $\rho = \psi^2$.

$$\psi(0) = 0.9701330,$$

$$I_1 = \int (\nabla\psi)^2 = 8.5838197, \quad (7.25)$$

$$I_2 = \int \{r^{-5/2} - \rho^{5/3}\} = 42.92,$$

$$I_3 = \int [r^{-3/2} - \rho]/r = 34.34.$$

From Eq. (7.17) one has $I_1 + I_3 = I_2$. By dilating $\psi(r) \rightarrow t^{3/4}\psi(tr)$ in Eq. (7.19), a "virial theorem" is obtained: $5I_1 + 3I_2 = 5I_3$. Thus

$$I_1 : I_2 : I_3 = 1 : 5 : 4, \quad (7.26)$$

$$\Delta E = \mathcal{F}(\psi) = I_1 - 3I_2/5 + I_3 = 2I_1. \quad (7.27)$$

If the parameters are reintroduced

$$I_1(z, \gamma, A) = A \int (\nabla\psi)^2 = z^2 A^{1/2} \gamma^{-3/2} I_1. \quad (7.28)$$

Let us denote the ρ we have just obtained in Theorem 7.29 [with the parameters reintroduced according to Eq. (7.18)] by ρ_∞ . The scale length of ρ_∞ is z^{-1} and, for large r , $\rho_\infty(r)$ agrees to leading order with $\rho^{\text{TF}}(r)$ for small r (on a scale of $z^{-1/3}$), namely, $(z/\gamma r)^{3/2}$. We claim that ρ_∞ can be spliced together with ρ^{TF} in the overlap region, $r = O(z^{-2/3})$, and the result is ρ^{TFW} to leading order in z . The splicing is independent of λ provided $\lambda/z > (\text{const}) > 0$. The change in energy for an atom is then, to leading order, ΔE of Eq. (7.27), and is independent of λ . An analogous situation holds for a molecule; near each nucleus ρ^{TF} is spliced together with ρ_∞ for the appropriate z_j . This is formalized in the following theorem.

Theorem 7.30. Let $V(x) = \sum z_j |x - R_j|^{-1}$. Consider the $Z \rightarrow \infty$ limit with the scaling given before Eq. (5.2), except that the electron charge N is not restricted to be integral. $\lambda = N/Z > 0$ is fixed. $z_j = az_j^0$, $R_j = a^{-1/3}R_j^0$, with $a\lambda = N$. Then, as $N \rightarrow \infty$,

$$(i) E^{\text{TFW}}(N) = E^{\text{TF}}(N) + D \sum_{j=1}^k z_j^2 + O(a^2), \quad (7.29)$$

with $D = 2A^{1/2}\gamma^{-3/2}I_1$.

$$(ii) a^{-4/3}\mu^{\text{TFW}}(N) \rightarrow \mu^{\text{TF}}(\lambda, \underline{z}^0, \underline{R}^0). \quad (7.30)$$

(iii) Fix x . Then

$$a^{-2}\rho^{\text{TFW}}(N, \underline{z}, R; a^{-1/3}x) \rightarrow \rho^{\text{TF}}(\lambda, \underline{z}^0, R^0; x), \quad (7.31)$$

with convergence in the sense of weakly in L^1 if $\lambda \leq Z$ and weakly in L^1_{loc} if $\lambda > Z$.

(iv) Fix y . For each j

$$z_j^{-3}\rho^{\text{TFW}}(N, \underline{z}, R; z_j + z_j^{-1}y) \rightarrow (Ay)^{-3/2}\psi^2(y/A), \quad (7.32)$$

where ψ is the solution to Eq. (7.17) with $A = z = \gamma = 1$ given by Theorem 7.29. The convergence is pointwise and in L^1_{loc} . A refinement of Eq. (7.32) is given in Theorems 7.32-7.35.

Before proving Theorem 7.30 let us comment on its significance.

(i) Equation (7.29) states that the energy correction in TFW theory is exactly of the form of the quantum correction conjectured by Scott [Eq. (5.29)]. In particular, since $\gamma^{3/2} \sim q^{-1}$, the q dependence is the same. In order to obtain the conjectured coefficient $\frac{1}{8}$ of Eq. (5.32), with $\gamma = \gamma_\rho$, we must choose

$$A = q^2 \gamma_\rho^3 [16I_1]^{-2} = 0.18590919. \quad (7.33)$$

This number was mentioned after Eq. (2.8).

Yonei and Tomishima (1965) also realized that $A = 1/5$ is a good choice. They analyzed the TFW atom without electron repulsion, namely Eq. (7.16), and compared the

TFW energy with the quantum Bohr energy, Eq. (5.31), for neutral atoms with z up to 100. They did not seem to notice that this choice for A is valid even if $\lambda = N/z < 1$. Yonei (1971) analyzed TFDW theory with electron repulsion and again advocated $A = 1/5$. This is not surprising since Theorem 7.30 says that the electron repulsion does not affect ΔE to $O(z^2)$ and Theorem 6.5 (suitably modified) says that the Dirac correction changes the energy to $O(z^{5/3})$. Yonei (1971) claims that the dissociation energy and the equilibrium internuclear distance for the nitrogen molecule, calculated with this TFDW theory, are in good agreement with experiment.

(ii) The density, on a length scale $z^{-1/3}$ agrees with quantum (and TF) theory, Theorem 5.2.

(iii) On a length scale z^{-1} near each nucleus, Eq. (7.32) states that ρ^{TFW} converges to a universal function. This phenomenon is the same as we conjectured in Eq. (5.37) for quantum theory. The universal functions are not exactly the same, but they are very close. For large values of the argument they agree, namely, $(\gamma_p y)^{-3/2}$, independent of A . Since the convergence in Eq. (7.32) is *pointwise*, it makes sense to ask what happens at $y = 0$. Using γ_p and A given by Eq. (7.33), the right side of (7.32) is obtained from (7.18) and (7.25) as

$$q^{-1} z_j^{-3} \rho^{\text{TFW}}(x = R_j) \rightarrow 0.198\,271\,49. \quad (7.34)$$

On the other hand, ρ^n in Eq. (5.33) can be evaluated at $x = 0$, since only S waves contribute. At $x = 0$, $f_{\text{noo}}(0)^2 = (8\pi n^2)^{-1}$. Thus Eq. (5.37), if correct, would state that

$$q^{-1} z_j^{-3} \rho^0(x = R_j) \rightarrow \zeta(3)/8\pi = 0.047\,828\,325. \quad (7.35)$$

To prove Theorem 7.30, Theorems 7.32–7.35, which are independently interesting, are needed. To prove them we need the following comparison theorem which was proved by Morgan (1978) in the spherically symmetric case and by T. Hoffmann-Ostenhof (1980) in the general case.

Lemma 7.31. *Let $B \subset \mathbb{R}^3$ be open, and let f and g be continuous functions on the closure of B that satisfy Δf and $\Delta g \in L^1(B)$ and $f(x)$ and $g(x) \rightarrow 0$ as $|x| \rightarrow \infty$ if B is unbounded. Assume $\Delta f \leq Ff$ and $\Delta g \geq Gg$ as distributions on B , where F, G are functions satisfying $F(x) < G(x)$ a.e. in B . Assume $f(x) > 0$ in B and $f(x) \geq g(x)$ for all $x \in \partial B$. Then $f(x) \geq g(x)$ for all $x \in B$.*

Theorem 7.32. *Let $V = z/|x|$, and let ψ_μ be the positive solution to Eq. (7.17) given in Theorem (7.29). Let ψ be the positive solution to the TFW equation, (7.9), for some $\mu \geq 0$ and $p = \frac{5}{3}$. Then, for all x , $\psi(x) \leq \psi_\mu(x)$.*

Proof. Let $B = \{x \mid \psi(x) - \psi_\mu(x) > 0\}$. Take $f = \psi_\mu$ and $g = \psi$ in Lemma 7.31. Since f and g are continuous and, by Theorem 7.24, $g(x) < f(x)$ for large $|x|$, B is open and bounded. Hence Δf and $\Delta g \in L^1(B)$. On B , $A\Delta f = Ff$, $A\Delta g > Gg$ with $F = -V + f^{4/3}\gamma$ and $G = -V + g^{4/3}\gamma$. Since $F < G$ in B and $f = g$ on ∂B , $f \geq g$ in B . Therefore, B is empty. ■

For a molecule, an upper bound to ψ , which is not as nice as Theorem 7.32 but which is sufficient for Theorem 7.30, can also be obtained. We always assume p

$$= \frac{5}{3}.$$

Theorem 7.33. *Let V be as in Theorem 7.30 with the scaling given there. Let ψ be the positive solution to the TFW equation for $\mu \geq 0$ and let B be the ball $\{x \mid |x - R_1| < r\}$. Then, for sufficiently large a ,*

$$\psi(x) \leq \psi_\mu(x - R_1) \quad \text{for } x \in B \quad (7.36)$$

where ψ_μ is the positive solution to Eq. (7.17) with

$$z = z_1 + dZ^{2/3}$$

and

$$d = 1 + 2(Z^0)^{-1/3} / \min \{ |R_j^0 - R_1^0| \mid j = 2, \dots, k \}.$$

Proof. By Theorem 7.14 and Eq. (7.24), we can choose a large enough so that $\psi_\mu(x - R_1) > \psi(x)$ when $x \in \partial B$ and so that $R_2, \dots, R_k \notin B$. The proof is then the same as for Theorem 7.32, with $f = \psi_\mu$ and $g = \psi$, provided we can verify that $M(x) = z|x - R_1|^{-1} - V(x) > 0$ when $x \in B$. But M , being superharmonic in B , has its minimum on ∂B . This minimum is positive for large enough a . ■

To obtain a lower bound to ψ , the following is needed.

Theorem 7.34. *Assume the hypothesis of Theorem 7.30 with $\lambda > 0$ and let ψ be the positive solution to the TFW equation. Then there is a constant d , independent of λ , such that*

- (i) $h(x) \equiv |x|^{-1} \star \rho < da^{4/3}$.
- (ii) $\mu < da^{1/3} \lambda^{-2/3}$.

Proof. For (i) we use Theorem 7.14 together with the fact that $\int \rho = a\lambda$. For any x , let B be the ball of radius $a^{-1/3}$ centered at x . The contribution to h from $\chi_B \rho$ is bounded by $(\text{const})(Z^0)^{3/2} a^{4/3}$. The contribution from $(1 - \chi_B)\rho$ is bounded by $a^{1/3} \int \rho$. For (ii), since μ is decreasing in λ , $\mu \leq -e(N)/N$. However, $e(N) > e^{\text{TF}}(N)$. But $-e^{\text{TF}}(N)$ scales as $a^{1/3} f(\lambda)$ and $f(\lambda) \leq (\text{const})\lambda^{1/3}$ by Eq. (3.6). ■

Theorem 7.35. *Assume the same hypothesis as in Theorem 7.33. Then, for sufficiently large a ,*

$$\psi(x) \geq \psi_\mu(x - R_1)\sigma(x - R_1) \quad \text{for all } x, \quad (7.37)$$

where ψ_μ is the positive solution to Eq. (7.17) with $z = z_1 - 4ta^{2/3}A$ and

$$\sigma(x) = [1 - a^{2/3}t|x|] \exp(-a^{2/3}t|x|).$$

Here, $At^2 = d(1 + \lambda^{-2/3})$ with d given in Theorem 7.34.

Proof. Let $f = \psi$ and $g =$ right side of Eq. (7.37). We have to verify (7.37) only in $B = \{x \mid a^{2/3}t|x - R_1| < 1\}$ because $g \leq 0$ otherwise. Since, by Theorem 7.29, both ψ_μ and σ are symmetric decreasing, $\Delta g > \sigma \Delta \psi_\mu + \psi_\mu \Delta \sigma$. But

$$(\Delta \sigma)(x) = (a^{4/3}t^2 - 4a^{2/3}t/|x|)\sigma,$$

and $\psi_\mu^{4/3} \geq g^{4/3}$ since $\sigma \leq 1$. Therefore, to imitate the proof of Theorem 7.32, it is only necessary to verify that $a^{4/3}At^2 > h(x) + \mu$, but this is clearly true. ■

Proof of Theorem 7.30. (iv) is a trivial consequence of Theorems 7.33 and 7.35. (iii) is proved in the same way as Theorem 5.2 if we note that the energy can be controlled to $O(z^{1/3})$ by the variational upper bound given in the paragraph after Eq. (7.15). (ii) is proved by noting

that, by the proof of (iii) just given,

$$a^{-7/3} E^{\text{TFW}}(a, N = a\lambda) \rightarrow E^{\text{TF}}(a = 1, \lambda).$$

The limit of the derivative of a sequence of convex (in λ) functions is the derivative of the limit function.

The proof of (i) is complicated. Upper and lower bounds to E of the desired accuracy, $O(Z^2)$, are needed. First, let us make a remark. Consider E as a function of A . By standard arguments used earlier, $E(A)$ is monotone increasing, concave, and hence differentiable almost everywhere for $A > 0$. $dE/dA = T/A$ a.e., and $E^{\text{TFW}} - E^{\text{TF}} = \int_0^A (T/A) dA$. If we can find a lower and upper bound to T/A of the form

$$T/A = A^{-1/2} \gamma^{-3/2} \sum z_j^2 + (\text{lower order})$$

then Eq. (7.29) will be proved. We can, indeed, find a lower bound of this form, and hence a lower bound to E . We cannot find an upper bound of this form and therefore must resort to a direct variational calculation to obtain an upper bound to E .

Upper bound. By the monotonicity of E in N , it is only necessary that $\int \rho \leq N$. There are several ways to construct a variational ρ , which we call f . The details of the calculation of $\mathcal{E}(f)$ are left to the reader. One construction is to define $B = \{x | \rho^{\text{TF}}(x) > Z^{5/2}\}$. For large a , B is the union of k connected components which are approximately spheres centered at R_j . Call these B_j . Let ψ_{ω_j} be the solution to Eq. (7.17) centered at R_j and with $z = z_j - ta^{2/3}$. Let $C_j = \{x | \psi_{\omega_j} > Z^{5/4}\}$. For large enough, but fixed t , $C_j \subset B_j$ for large a . The variational f is defined by $f(x) = \rho^{\text{TF}}(x)$ for $x \notin B$, $f(x) = Z^{5/2}$ for $x \in B_j \setminus C_j$, and $f(x) = \psi_{\omega_j}(x)^2$ for $x \in C_j$.

Lower bound. We construct a lower bound to T/A . Suppose $P_1, \dots, P_k$ are orthogonal, vector valued functions. Then $T/A \geq \sum_j L_j^2 / \int P_j^2$, where $L_j = \int \nabla \psi \cdot P_j$. We take $P_j(x) = \nabla \psi_{\omega_j}(x) \chi_j(x)$, where χ_j is the characteristic function of $D_j = \{x | |x - R_j| < tz_j^{-2/3}\}$, and t is some fixed constant. For large a , the D_j are disjoint so the P_j are orthogonal. Clearly, $\int P_j^2 = \int \nabla \psi_{\omega_j}^2 + o(Z^2)$. Now multiply Eq. (7.17) for ψ_{ω_j} by ψ and integrate over D_j . Then

$$L_j = -A^{-1} \int W_{\omega_j} \psi_{\omega_j} \psi \chi_j + \int \psi \nabla \psi_{\omega_j} \cdot n ds.$$

By the bound (7.37), the first integral is $(T_{\omega_j}/A) + o(Z^2)$. It is not difficult to show that the second integral is $o(Z^2)$. This can be done by using Eq. (7.24), whence, for some $t \in [\frac{1}{2}, 1]$, $d\psi_{\omega_j}/dr > -10z_j^{3/4} r^{-7/4}$ at $r = tz_j^{-2/3}$. ■

VIII. THOMAS-FERMI-DIRAC-VON WEIZSÄCKER THEORY

This theory has not been as extensively studied as the other theories. The results presented here are from unpublished work by Benguria, Brezis, and Lieb done in connection with their 1981 paper.

The energy functional is

$$\begin{aligned} \mathcal{E}'(\psi) = & A \int (\nabla \psi)^2 + \int J(\psi^2) \\ & - \int V \psi^2 + D(\psi^2, \psi^2) + U \end{aligned} \quad (8.1)$$

in units in which $\hbar^2/2m = 1$.

$$J(\rho) = (\gamma/\rho) \rho^p - (3C_e/4) \rho^{4/3}. \quad (8.2)$$

For convenience we assume $p > \frac{4}{3}$ (not $p > 1$). $\mathcal{E}(\rho) \equiv \mathcal{E}'(\rho^{1/2})$. The function space for ψ is the same as for TFW theory, namely, G'_0 of Eq. (7.3). Note that $\mathcal{E}(\rho)$ is not convex because of the $-\int \rho^{4/3}$ term.

As in TFD theory Eqs. (6.7)–(6.10), we introduce

$$J_\alpha(\rho) = J(\rho) + \alpha \rho, \quad (8.3)$$

and α is chosen so that $J_\alpha(\rho) \geq 0$ and $J_\alpha(\rho_0) = 0 = J'_\alpha(\rho_0)$ for some ρ_0 , namely,

$$\begin{aligned} \rho_0^{p-4/3} &= C_e \rho [4\gamma(\rho-1)]^{-1} \\ \alpha &= (3p-4)[4(\rho-1)]^{-1} \rho_0^{1/3} C_e. \end{aligned} \quad (8.4)$$

The necessity of $p > \frac{4}{3}$ for this construction is obvious. $\mathcal{E}'_\alpha$ and $\mathcal{E}_\alpha$ are defined by using J_α in Eq. (8.1).

The energy for $\lambda \geq 0$ is

$$E(\lambda) = \inf \left\{ \mathcal{E}(\rho) \mid \rho \in G_p, \int \rho = \lambda \right\}, \quad (8.5)$$

and similarly for $E_\alpha(\lambda)$ and $E'(\lambda)$, $E'_\alpha(\lambda)$ using $\mathcal{E}'$. If the condition $\int \rho = \lambda$ is omitted in (8.5) we obtain E , E_α , E' , E'_α .

Theorem 8.1. (i) The four functions $E(\lambda)$, $E_\alpha(\lambda)$, $E'(\lambda)$, and $E'_\alpha(\lambda)$ are finite, continuous, and satisfy

$$E(\lambda) = E'(\lambda) = E_\alpha(\lambda) - \alpha \lambda = E'_\alpha(\lambda) - \alpha \lambda. \quad (8.6)$$

(ii) E_α is finite.

(iii) ρ minimizes $\mathcal{E}(\rho)$ on $\int \rho = \lambda$ if and only if $\psi = \rho^{1/2}$ minimizes $\mathcal{E}'(\psi)$ on $\int \psi^2 = \lambda$. This ρ and ψ also obviously minimize $\mathcal{E}_\alpha$ and $\mathcal{E}'_\alpha$.

Proof. The same as for Theorems 2.1, 6.2, and 7.2. Note that $[\int \rho^{4/3}]^{3p-3} \leq [\int \rho]^{3p-4} \int \rho^p$ (by Hölder). ■

Theorem 8.2. Let ψ minimize $\mathcal{E}'(\psi)$ on the set $\int \psi^2 = \lambda$. Then ψ satisfies the TFDW equation:

$$[-\Delta + W(x)]\psi = -\mu\psi, \quad (8.7)$$

in the sense of distributions, with

$$W = \gamma \rho^{p-1} - C_e \rho^{1/3} - \phi + \alpha, \quad (8.8)$$

$\phi = V - |x|^{-1} \star \rho$, and $\rho = \psi^2$. Apart from a sign, $\psi(x) > 0$ for all x , and ψ satisfies the conclusions of Theorem 7.9. ψ is the unique ground state of $H = -\Delta + W(x)$ and μ is its ground-state eigenvalue. E is differentiable at λ and $\mu = -dE_\alpha/d\lambda = -dE/d\lambda - \alpha \geq -\alpha$. $\mu = 0$ if $E_\alpha(\lambda)$ has an absolute minimum at this λ .

Proof. The proof is basically the same as for Theorems 7.8–7.10. Although it is not known that $\rho = \psi^2$ is unique, this is not really necessary. By considering the variation of $\mathcal{E}'(\psi)$, ψ satisfies Eqs. (8.7) and (8.8). If ψ is minimizing, then so is $|\psi|$ (cf. Theorem 7.2). Hence $|\psi|$ satisfies Eq. (8.7) with the same W . But, as in Theorem 7.10, the ground state of $H = -\Delta + W$ is unique and non-negative and therefore ψ may be taken to be ≥ 0 for all x . The rest follows by the methods of Theorem 7.9. (Note: $\rho^{1/3} \psi \in L^3$ since $\psi \in L^6 \cap L^2$). ■

Remark. As in Sec. VII, the role of $\mathcal{E}'$, as distinct from $\mathcal{E}$, is solely to prove Eq. (8.7), in which no explicit reference to $\rho \geq 0$ is made.

Remark. Theorem 8.2 does not assert the existence of

a minimizing ψ with $\int \psi^2 = \lambda$.

Now we turn to a difficult and serious problem. We do not know that $E_\alpha(\lambda)$ is monotone nonincreasing. Therefore, if we define

$$\hat{E}_\alpha(\lambda) = \inf \left\{ \mathcal{E}_\alpha(\rho) \mid \rho \in G_\rho, \int \rho \leq \lambda \right\}, \quad (8.9)$$

we do not know that $E_\alpha(\lambda) = \hat{E}_\alpha(\lambda)$. By definition, $\hat{E}_\alpha(\lambda)$ is monotone nonincreasing. The source of the difficulty is this: Although $J_\alpha(\rho_0) = J'_\alpha(\rho_0) = 0$ (as in TFD theory), we cannot simply add small clumps of charge, of amplitude ρ_0 , at ∞ . This is so because such a clump would then have $\int (\nabla \psi)^2 = \infty$. Nevertheless, we can add clumps with $\mathcal{E}_\alpha$ energy strictly less than $\alpha \int \rho$, as the following theorem shows.

Theorem 8.3. *Set $V=0$ in $\mathcal{S}'$. There are C^∞ functions of compact support such that $\mathcal{S}'(\psi) < 0$.*

Proof. Let f be any function in C_0^∞ and let $\psi(x) = b^2 f(bx)$. For some sufficiently small b , but positive b , $\mathcal{S}'(\psi) < 0$. To see this, note that $\int (\nabla \psi)^2$ scales as b^3 , $\int \rho^{5/3}$ as $b^{11/3}$, $D(\rho, \rho)$ as b^3 , while $\int \rho^{4/3}$ scales as $b^{7/3}$. ■

As a corollary we have the following.

Theorem 8.4. *$E(\lambda)$ is strictly monotone decreasing in λ . Hence*

$$E(\lambda) = \inf \left\{ \mathcal{E}(\rho) \mid \rho \in G_\rho, \int \rho \leq \lambda \right\}. \quad (8.10)$$

$$E = \inf_\lambda E(\lambda) = -\infty.$$

We conjecture that $E(\lambda)$ is convex. Unfortunately, the "convexification" trick of Sec. VI, in which J_α is replaced by j , is not helpful. Because of the gradient term, any minimizing ψ will be continuous, and therefore ψ cannot omit the values $(0, \rho_0)$, even for point nuclei. While the energy for j is, indeed, convex, it is strictly smaller than $E_\alpha(\lambda)$ for all λ .

Theorem 8.5 states that E_α and $\hat{E}_\alpha$ have absolute minima at some common, finite λ . For all we know, there may be several such λ , but all these λ are bounded. Furthermore, for every λ there is a minimizing ρ for $\hat{E}_\alpha(\lambda)$. Unfortunately, for no λ are we able to infer that $\int \rho \leq \lambda$.

Theorem 8.5. (i) *There exists a minimizing ρ for $\mathcal{E}_\alpha(\rho)$ on G_ρ , and $\psi = \rho^{1/2}$ minimizes $\mathcal{S}'_\alpha(\psi)$. Every such $\rho \in L^1$, and $\int \rho \leq$ some constant which is independent of ρ .*

(ii) *There exists a minimizing ρ for $\mathcal{E}_\alpha(\rho)$ on the set $\int \rho \leq \lambda$.*

Remark. It is not claimed (but it is conjectured) that the minimizing ρ is unique.

Proof. The proofs of (i) and (ii) are the same, so we concentrate on (i). The proof merely imitates the proof of Theorem 7.6. The only new point is that $\rho \in L^1$. Each term in $\mathcal{E}(\rho)$ is finite and, in particular, $I = \int J_\alpha(\rho) < \infty$. But $J_\alpha(\rho) \geq k\rho$ when $0 \leq \rho \leq \beta$ for some $k, \beta > 0$. If χ is the characteristic function of $\{x \mid \rho(x) \leq \beta\}$, $k \int \chi \rho < \infty$. On the other hand, $\beta^2(1-\chi)\rho \leq \rho^3$, so $\beta^2 \int (1-\chi)\rho \leq \int \rho^3 < \infty$ since $\rho \in L^3$. It is easy to see from Eq. (7.6) that the bound on $\int \rho$ is independent of ρ . ■

Remark. It is surprising that the fact that $\lambda < \infty$ for

any absolute minimum is obtained so easily. Recall that in TFW theory the proof of this fact (Theorem 7.12) required analysis of the TFW equation.

An important question is whether λ , for an absolute minimum, always satisfies $\lambda \geq Z$. A few things can be said about the properties of any minimizing ρ on $\int \rho = \lambda$.

Theorem 8.6. *In the atomic case, $V(r) = z/r$, any minimizing ψ is symmetric decreasing when $\lambda \leq z$. (Conjecture: this also holds for all λ .)*

Proof. The rearrangement inequality proof of Theorem 2.12 is applicable. ■

Theorem 8.7. *The conclusions of Theorem 7.13 hold for any minimizing ψ . Moreover, for every $t < \mu + \alpha$ there exists a constant M such that*

$$\psi(x) \leq M \exp[-(t/A)^{1/2}|x|].$$

Proof. Same as for Theorems 7.13 and 7.24. ■

Theorem 8.8. *Every minimizing ψ satisfies Theorem 7.25.*

Plainly, TFDW theory is not in a satisfactory state from the mathematical point of view. In TFD theory we were able to deal with the lack of convexity by means of the J_α trick. In TFW theory, the presence of the gradient term does not spoil the general theory because $\mathcal{S}$ is convex. When taken together, however, the two difficulties present an unsolved mathematical problem.

ACKNOWLEDGMENTS

I am grateful to the U. S. National Science Foundation (Grant No. PHY-7825390-A02) for supporting this work. Thanks go to Freeman Dyson and Barry Simon for a critical reading of the manuscript.

REFERENCES

- Adams, R. A., 1975, *Sobolev Spaces* (Academic, New York).
- Balázs, N., 1967, "Formation of stable molecules within the statistical theory of atoms," *Phys. Rev.* **156**, 42–47.
- Baumgartner, B., 1976, "The Thomas–Fermi theory as result of a strong-coupling limit," *Commun. Math. Phys.* **47**, 215–219.
- Baxter, J. R., 1980, "Inequalities for potentials of particle systems," *Ill. J. Math.* **24**, 645–652.
- Benguria, R., 1979, "The von Weizsäcker and exchange corrections in Thomas–Fermi theory," Ph.D. thesis, Princeton University (unpublished).
- Benguria, R., 1981, "Dependence of the Thomas–Fermi Energy on the Nuclear Coordinates," *Commun. Math. Phys.*, to appear.
- Benguria, R., H. Brezis, and E. H. Lieb, 1981, "The Thomas–Fermi–von Weizsäcker theory of atoms and molecules," *Commun. Math. Phys.* **79**, 167–180.
- Benguria, R., and E. H. Lieb, 1978a, "Many-body potentials in Thomas–Fermi theory," *Ann. of Phys. (N.Y.)* **110**, 34–45.
- Benguria, R., and E. H. Lieb, 1978b, "The positivity of the pressure in Thomas–Fermi theory," *Commun. Math. Phys.* **63**, 193–218, Errata **71**, 94 (1980).
- Berestycki, H., and P. L. Lions, 1980, "Existence of stationary states in nonlinear scalar field equations," in *Bifurcation Phenomena in Mathematical Physics and Related Topics*, edited by C. Bardos and D. Bessis (Reidel, Dordrecht), 269–292. See also "Nonlinear Scalar field equations, Parts I and II," *Arch. Rat. Mech. Anal.*, 1981, to appear.
- Brezis, H., 1978, "Nonlinear problems related to the Thomas–Fermi equation," in *Contemporary Developments in Continuum Mechanics and Partial Differential Equations*, edited by

- G. M. de la Penha, and L. A. Medeiros (North-Holland, Amsterdam), 81-89.
- Brezis, H., 1980, "Some variational problems of the Thomas-Fermi type," in *Variational Inequalities and Complementarity Problems: Theory and Applications*, edited by R. W. Cottle, F. Giannessi, and J.-L. Lions (Wiley, New York), 53-73.
- Brezis, H., and E. H. Lieb, 1979, "Long range atomic potentials in Thomas-Fermi theory," *Commun. Math. Phys.* **65**, 231-246.
- Brezis, H., and L. Veron, 1980, "Removable singularities of nonlinear elliptic equations," *Arch. Rat. Mech. Anal.* **75**, 1-6.
- Caffarelli, L. A., and A. Friedman, 1979, "The free boundary in the Thomas-Fermi atomic model," *J. Diff. Equ.* **32**, 335-356.
- Deift, P., W. Hunziker, B. Simon, and E. Vock, 1978, "Pointwise bounds on eigenfunctions and wave packets in N -body quantum systems IV," *Commun. Math. Phys.* **64**, 1-34.
- Dirac, P. A. M., 1930, "Note on exchange phenomena in the Thomas-Fermi atom," *Proc. Cambridge Philos. Soc.* **26**, 376-385.
- Fermi, E., 1927, "Un metodo statistico per la determinazione di alcune prioriet  dell'atome," *Rend. Accad. Naz. Lincei* **6**, 602-607.
- Firsov, O. B., 1957, "Calculation of the interaction potential of atoms for small nuclear separations," *Zh. Eksp. i Teor. Fiz.* **32**, 1464. [English transl. *Sov. Phys.—JETP* **5**, 1192-1196 (1957)]. See also *Zh. Eksp. Teor. Fiz.* **33**, 696 (1957); **34**, 447 (1958) [*Sov. Phys.—JETP* **6**, 534-537 (1958); **7**, 308-311 (1958)].
- Fock, V., 1932, " ber die G ltigkeit des Virialsatzes in der Fermi-Thomas'schen Theorie," *Phys. Z. Sowjetunion* **1**, 747-755.
- Gilbarg, D., and N. Trudinger, 1977, *Elliptic Partial Differential Equations of Second Order* (Springer Verlag, Heidelberg).
- Gombas, P., 1949, *Die statistischen Theorie des Atomes und ihre Anwendungen* (Springer Verlag, Berlin).
- Hille, E., 1969, "On the Thomas-Fermi equation," *Proc. Nat. Acad. Sci. (USA)* **62**, 7-10.
- Hoffmann-Ostenhof, M., T. Hoffmann-Ostenhof, R. Ahlrichs, and J. Morgan, 1980, "On the exponential falloff of wave functions and electron densities," *Mathematical Problems in Theoretical Physics, Proceedings of the International Conference on Mathematical Physics held in Lausanne, Switzerland, August 20-25, 1979*, Springer Lectures Notes in Physics, edited by K. Osterwalder (Springer-Verlag, Berlin, Heidelberg, New York, 1980), Vol. 116, 62-67.
- Hoffmann-Ostenhof, T., 1980, "A comparison theorem for differential inequalities with applications in quantum mechanics," *J. Phys. A* **13**, 417-424.
- Jensen, H., 1933, " ber die G ltigkeit des Virialsatzes in der Thomas-Fermischen Theorie," *Z. Phys.* **81**, 611-624.
- Kato, T., 1957, "On the eigenfunctions of many-particle systems in quantum mechanics," *Commun. Pure Appl. Math.* **10**, 151-171.
- Lee, C. E., C. L. Longmire, and M. N. Rosenbluth, 1974, "Thomas-Fermi calculation of potential between atoms," Los Alamos Scientific Laboratory Report No. LA-5694-MS.
- Lieberman, D. A., and E. H. Lieb, 1981, "Numerical calculation of the Thomas-Fermi-von Weizs cker function for an infinite atom without electron repulsion," Los Alamos National Laboratory Report in preparation.
- Lieb, E. H., 1974, "Thomas-Fermi and Hartree-Fock theory," in *Proceedings of the International Congress of Mathematicians, Vancouver*, Vol. 2, 383-386.
- Lieb, E. H., 1976, "The stability of matter," *Rev. Mod. Phys.* **48**, 553-569.
- Lieb, E. H., 1977, "Existence and uniqueness of the minimizing solution of Choquard's nonlinear equation," *Stud. in Appl. Math.* **57**, 93-105.
- Lieb, E. H., 1979, "A lower bound for Coulomb energies," *Phys. Lett. A* **70**, 444-446.
- Lieb, E. H., 1981a, "A variational principle for many-fermion systems," *Phys. Rev. Lett.* **46**, 457-459; Erratum **47**, 69 (1981).
- Lieb, E. H., 1981b, "Analysis of the Thomas-Fermi-von Weizs cker equation for an atom without electron repulsion," in preparation.
- Lieb, E. H., and S. Oxford, 1981, "An improved lower bound on the indirect Coulomb energy," *Int. J. Quantum Chem.* **19**, 427-439.
- Lieb, E. H., and B. Simon, 1977, "The Thomas-Fermi theory of atoms, molecules and solids," *Adv. in Math.* **23**, 22-116. These results were first announced in "Thomas-Fermi theory revisited," *Phys. Rev. Lett.* **31**, 681-683 (1973). An outline of the proofs was given in Lieb, 1974.
- Lieb, E. H., and B. Simon, 1978, "Monotonicity of the electronic contribution to the Born-Oppenheimer energy," *J. Phys. B* **11**, L537-542.
- Lieb, E. H., and W. Thirring, 1975, "Bound for the kinetic energy of fermions which proves the stability of matter," *Phys. Rev. Lett.* **35**, 687-689; Errata **35**, 1116 (1975).
- Lieb, E. H., and W. Thirring, 1976, "A bound for the moments of the eigenvalues of the Schroedinger Hamiltonian and their relation to Sobolev inequalities," in *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*, edited by E. H. Lieb, B. Simon, and A. S. Wightman (Princeton University Press, Princeton), 269-303.
- March, N. H., 1957, "The Thomas-Fermi approximation in quantum mechanics," *Adv. in Phys.* **6**, 1-98.
- Mazur, S., 1933, " ber konvexe Mengen in linearen normierten R umen," *Studia Math.* **4**, 70-84. See p. 81.
- Morgan, J., III., 1978, "The asymptotic behavior of bound eigenfunctions of Hamiltonians of single variable systems," *J. Math. Phys.* **19**, 1658-1661.
- Morrey, C. B., Jr., 1966, *Multiple integrals in the calculus of variations* (Springer, New York).
- O'Connor, A. J., 1973, "Exponential decay of bound state wave functions," *Commun. Math. Phys.* **32**, 319-340.
- Reed, M., and B. Simon, 1978, *Methods of Modern Mathematical Physics* (Academic, New York), Vol. 4.
- Ruskai, M. B., 1981, "Absence of discrete spectrum in highly negative ions," *Commun. Math. Phys.* (to appear).
- Scott, J. M. C., 1952, "The binding energy of the Thomas-Fermi atom," *Philos. Mag.* **43**, 859-867.
- Sheldon, J. W., 1955, "Use of the statistical field assumption in molecular physics," *Phys. Rev.* **99**, 1291-1301.
- Simon, B., 1981, "Large time behavior of the L^p norm of Schroedinger semigroups," *J. Func. Anal.* **40**, 66-83.
- Stampacchia, G., 1965, *Equations elliptiques du second ordre a coefficients discontinus* (Presses de l'Universite, Montreal).
- Teller, E., 1962, "On the stability of molecules in the Thomas-Fermi theory," *Rev. Mod. Phys.* **34**, 627-631.
- Thirring, W., 1981, "A lower bound with the best possible constant for Coulomb Hamiltonians," *Commun. Math. Phys.* **79**, 1-7 (1981).
- Thomas, L. H., 1927, "The calculation of atomic fields," *Proc. Camb. Philos. Soc.* **23**, 542-548.
- Torrens, I. M., 1972, *Interatomic Potentials* (Academic, New York).
- Veron, L., 1979, "Solutions singuli res d'equations elliptiques semilinaires," *C. R. Acad. Sci. Paris* **288**, 867-869. This is an announcement; details will appear in "Singular solutions of nonlinear elliptic equations," *J. Non-Lin. Anal.*, in press.
- von Weizs cker, C. F., 1935, "Zur Theorie der Kernmassen," *Z. Phys.* **96**, 431-458.
- Yonei, K., and Y. Tomishima, 1965, "On the Weizs cker correction to the Thomas-Fermi theory of the atom," *Jour. Phys. Soc. Japan* **20**, 1051-1057.
- Yonei, K., 1971, "An extended Thomas-Fermi-Dirac theory for diatomic molecules," *Jour. Phys. Soc. Japan* **31**, 882-894.

INDEX

L/R refer to left/right column

- admissible density matrix, 621L
- amenable potential, 625R, 627L
- asymptotics, 607L, 607R, 613R, 633R, 639R
- atomic radius, 614L, 624R
- atomic surface (see surface, atomic)
- Baxter's theorem, 615L
- Banach-Alaoglu theorem, 605R, 629R
- Benguria's theorem, 617R
- binding, 604L, 614R–615L, 628R, 634L–635R
- Bohr atom, 623L, 623R
- boundary, 607L
- chemical potential, 606R, 613L, 631L
- chemical potential (asymptotics), 613R, 631L
- compressibility, 617L
- convexification, 605L, 625R
- convexity, 605L
- Coulomb potential, 604L, 612R, 625L
- core, atomic, 624L
- critical density, 606L, 612R, 626L, 631L–633R
- density, 604L
- density matrix, admissible single particle, 621L
- dilation, 616R–618L, 619R
- dipole-dipole interaction, 620L
- Dirac correction, 604R, 625L
- domain of the energy functional, 604R, 625L, 628R, 636L
- double layer, 609R
- electronic contribution, 605L, 616L
- electron number, 604R, 620R
- energy functional, 604L, 604R, 611L, 625L, 628R, 638L
- exponential falloff, 604L, 624R, 633R, 639R
- Fatou's lemma, 605R, 629R
- Firsov's principle, 610L, 610R
- free boundary problem, 607L
- ground state, approximate, 623L, 624L
- harmonic, 607L
- heavy atom, 624L, 624R
- inner core, 624L
- infinite atom, 623R, 636L
- ions, negative, 604L, 628R, 632R, 634L
- ionization potential, 613R, 624R
- ionization, spontaneous, 634R
- j model, 614L, 625R
- kinetic energy, 608L, 617L, 625L
- L^p space, 605L
- long range interaction, 619R, 620L
- many-body potentials, 615R, 619R
- Mazur's theorem, 605R, 615L, 626L
- minimization, 605L, 606L, 625L, 628R, 638R
- minimizing density, 606L, 626L, 629R, 639L
- no-binding theorem, 603R, 611L, 614R
- nuclear charge, 604L
- nuclear coordinates, 604L
- nuclear potential, 604L
- nuclear repulsion, 604L
- outer shell, 624L
- over-screening, 609R
- periodic Coulomb potential, 609L
- periodic Thomas-Fermi equation, 609L
- potential theory, 606R, Sec. III
- pressure, 608R, 617L
- quantum theory, 603R, 617L, 620R
- repulsive electrostatic energy, 604L, 621R, 625L
- scaling, 608R, 610L, 619R, 620R
- Scott correction, 603R, 623L–624L, 636R
- screening, 609R, 627L
- singularities, 607R, 612L, 620L
- Sobolev inequality, 629L, 629R
- solids, 608R–609R
- spin state number (q), 604R, 620R, 636R
- strong singularity, 607R, 620L
- subadditive, 612R, 614R
- subharmonic, 607L
- superadditive, 614R, 618R, 619L
- superharmonic, 606R
- surface, atomic, 614L, 624R
- surface charge, 609R
- symmetric decreasing function, 608L, 634L, 639R
- Teller's lemma, 611L, 612L
- Teller's theorem, 611L, 614R
- Thomas-Fermi energy, 605L
- Thomas-Fermi equation, 606R
- Thomas-Fermi equation (generalized), 611L
- Thomas-Fermi differential equation, 607L
- Thomas-Fermi-Dirac equation, 611L, 626L
- Thomas-Fermi-von Weizsäcker equation, 630L
- Thomas-Fermi-Dirac-von Weizsäcker equation, 638R
- Thomas-Fermi potential, 606R–608R, 611L–612R, 616R
- total shielding, 627L
- under-screening, 627L
- variational principle, 608L, 613L
- Virial theorem, 608L, 608R
- von Weizsäcker correction, 604R, Sec. VII
- Z^2 correction in TFW theory (see Scott correction), 635R–638L
- $Z \rightarrow \infty$ limit, 620R

Erratum: Thomas-Fermi and related theories of atoms and molecules [Rev. Mod. Phys. 53, 603–641 (1981)]

Elliott H. Lieb

Department of Mathematics and Physics, Princeton University, POB 708, Princeton, New Jersey 08544

Please note the following corrections:

Page 604, line after Eq. (2.8): A an adjustable *positive* constant.

Page 606: The last equality in Eq. (2.18) should read $[\phi_p(x) - \mu]_+^*$. In the line after Eq. (2.18), $[]$ should read $[]_+^*$.

Page 620: In Eq. (5.1) change $+V(x)$ to $-V(x)$.

Page 621, line before Eq. (5.3): Change $+V(x)$ to $-V(x)$.

Page 623, line 7: $|\psi_N, H_N \psi_N\rangle$ should read $\langle \psi_N, H_N \psi_N \rangle$.

Page 623: Eq. (5.32) should read $D = q/8$.

*) To avoid further confusion this error has been corrected in this re-edition of the article on page 266 of this volume.

The Hartree-Fock Theory for Coulomb Systems

Elliott H. Lieb*

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08540, USA

Barry Simon**

Department of Physics, Yeshiva University, New York, New York 10033, USA

Abstract. For neutral atoms and molecules and positive ions and radicals, we prove the existence of solutions of the Hartree-Fock equations which minimize the Hartree-Fock energy. We establish some properties of the solutions including exponential falloff.

§ 1. Introduction

In this paper we discuss the Hartree (H) and Hartree-Fock (HF) theories associated with the purely Coulombic Hamiltonian of electrons interacting with static nuclei. Our purpose will be to prove that these theories exist (in the sense that the equations have solutions which minimize the H or HF energy) whenever the system has an excess positive charge after the removal of one electron. An announcement of these results was given in [22] and an outline of the proof was given in [19].

The precise quantum system is described by the Hamiltonian

$$H = - \sum_{i=1}^N \Delta_i + \sum_{i=1}^N V(x_i) + \sum_{i < j} |x_i - x_j|^{-1}, \quad (1)$$

where

$$V(x) = - \sum_{j=1}^k z_j |x - R_j|^{-1} \quad (2)$$

acting on the Hilbert space $\mathcal{H} = L_a^2(\mathbb{R}^{3N}; \mathbb{C}^{2N})$. We assume $z_j > 0$, all j . The subscript a on L^2 indicates that we are to consider functions in L^2 as $\Psi(x_1, \sigma_1; \dots; x_N, \sigma_N)$ with $x_i \in \mathbb{R}^3$, $\sigma_i \in \pm 1/2$ and only allow those Ψ antisymmetric under interchanges of i and j . The particles have two spin states, but we could allow q spin states in our analysis below with only notational changes. The physically correct Fermi statistics

* Research partially supported by U.S. National Science Foundation Grant MCS-75-21684

** Research partially supported by U.S. National Science Foundation under Grants MPS-75-11864 and MPS-75-20638. On leave from Departments of Mathematics and Physics, Princeton University, Princeton, NJ 08540, USA

(antisymmetric functions) which we impose turns out to be the most difficult; our method would apply equally well to any other kind of statistics.

In (2), the z_j are the charges of the nuclei at positions R_j . By a famous theorem of Kato [16], H is essentially self-adjoint on $C_0^\infty(\mathbb{R}^{3N}; \mathbb{C}^{2N})_a = \mathcal{D}_{\text{phys}}$, the C^∞ functions of compact support.

We set:

$$E_N^Q(z_i, R_i) \equiv \inf \{ (\Psi, H\Psi) \mid \Psi \in \mathcal{D}_{\text{phys}}; \|\Psi\| = 1 \} \quad (3)$$

which is defined to be the quantum ground state energy.

In 1928, Hartree [14] introduced an approximate method for finding E_N^Q . He apriori ignored the spin variables and the Pauli principle and considered product wave functions.

$$\Psi(x_1, \dots, x_N) = \prod_{i=1}^N u_i(x_i). \quad (4a)$$

Minimization of the functional

$$\mathcal{E}_H(u_1, \dots, u_N) = (\Psi, H\Psi) \quad (4b)$$

with the constraint $\|u_i\| = 1$ then leads to the Euler-Lagrange equation

$$\tilde{h}_i u_i = \varepsilon_i u_i, \quad (5a)$$

where the ε_i are Lagrange multipliers and

$$(\tilde{h}_i w)(x) = (-\Delta w)(x) + V(x)w(x) + R_\Psi^i(x)w(x), \quad (5b)$$

$$R_\Psi^i(x) = \sum_{j \neq i} \int |x-y|^{-1} |u_j(y)|^2 d^3 y. \quad (5c)$$

Note that in the H equations, (5), the h_i depend non-trivially on i . This is to be contrasted with the HF equations (7) where h is independent of i . Of course, the equations (5) formally only correspond to stationary points of $\mathcal{E}_H$ so there should be solutions corresponding to u 's that do not minimize $\mathcal{E}_H$. Hartree attempted to take the Pauli principle into account by seeking solutions with $u_1 = u_2$ and u_3 "approximately orthogonal to u_1 ", $u_3 = u_4$ etc. [We should also mention that Hartree's derivation of (5) did not go through a minimization in the variational principle—this is a refinement due to Slater [30] which led him to the HF equations.] A more systematic and satisfactory way to take the Pauli principle into account was discovered in 1930 independently by Fock [10] and Slater [30] yielding equations now called Hartree-Fock (HF) equations. One considers trial functions $u_i(x_i, \sigma_i)$; $i = 1, \dots, N$ with $(u_i, u_j) = \delta_{ij}$ and the Slater determinant

$$\Psi(x_1, \sigma_1, \dots, x_N, \sigma_N) = (N!)^{-1/2} \det(u_i(x_j, \sigma_j)) \quad (6a)$$

and minimizes

$$\mathcal{E}_{\text{HF}}(u_1, \dots, u_N) = (\Psi, H\Psi) \quad (6b)$$

with the constraint $(u_i, u_j) = \delta_{ij}$. The corresponding Euler-Lagrange equations are :

$$hu_i = \varepsilon_i u_i, \quad (7a)$$

$$(hw)(x) = (-\Delta w)(x) + V(x)w(x) + U_\Psi(x)w(x) - (K_\Psi w)(x), \quad (7b)$$

$$U_\Psi(x) = \sum_{j=1}^N \int |x-y|^{-1} |u_j(y)|^2 d^3 y, \quad (7c)$$

$$(K_\Psi w)(x) = \sum_{j=1}^N u_j(x) \int |x-y|^{-1} \overline{u_j(y)} w(y) d^3 y. \quad (7d)$$

U_Ψ is the “direct” interaction and K_Ψ is the “exchange” interaction. We will show that minimizing solutions of (7a) exist whenever $N < Z + 1$ where Z is the nuclear charge

$$Z = \sum_{j=1}^k z_j.$$

We make the convention that when u 's depending on spin are involved, as in (7c) and (7d) the symbol $\int d^3 y$ indicates also a sum over the spin variable attached to y . [We note that the naive Euler-Lagrange equations are more complicated than (7) but after a unitary change, $u_i^{\text{new}} = \sum a_{ij} u_j^{\text{old}}$ with a_{ij} a unitary $N \times N$ matrix, (7) results. The Slater determinant (6a) is unaffected by the change so that (6b) is unaffected. This is proved in Lemma 2.3 and is further discussed in many texts, e.g. Bethe-Jackiw [6]; it plays an important role in § 2 below.]

Irrespective of the physical content of the H and HF equations, (5), (7), it is far from evident that there exist any solutions of them, let alone minimizing solutions, for they are clearly complicated non-linear equations. Because the full N -body Schrödinger equation is, at present, virtually inaccessible to computer calculation while the HF equation, especially in the spherical approximation [6], is ideal for computer iterative solution, the HF equations are extensively used in quantum chemistry [27].

Before our work, the only existing theorems were for the Hartree equation (5) as follows: Reeken [26] considered the *restricted* Hartree equations for Helium, i.e. he considered (5) with $k=1$, $z_1=2$ and the additional restriction $u_1=u_2$. He found a solution for this case with $u_i \geq 0$ pointwise; his method works for any $z > 1$. Independently Gustafson and Sather [12] found solutions for the restricted two electron problems for sufficiently large z (they state their results for $z=2$ but with $\|u_i\|$ sufficiently small rather than 1. Since we insist on the normalization condition $\|u_i\|=1$, we scale coordinates to translate their result into a large z , $\|u_i\|=1$ result). These authors all use a bifurcation analysis further discussed in Stuart [32], and depend on the fact that they seek spherically symmetric solutions so that methods of ordinary differential equations are available. Properties of their solutions are further discussed in [3,4]. Relations between the restricted Hartree two electron problem and the unrestricted problem appear to present some interesting mathematical phenomena and we hope to return to them in a future publication.

Using a Schauder-Tychonoff theorem, Wolkowisky [34] found ground state and excited solutions of the Hartree equation in the spherical approximation (see e.g. Bethe-Jackiw [6] for a discussion of the approximation).

All these authors attack the equations directly as fixed point equations in some sense. The reason we are able to go further is that we exploit the form of the equations as gradient maps, i.e. as Euler-Lagrange equations and directly attempt to find solutions by finding minimizing u 's for $\mathcal{E}_H$ and $\mathcal{E}_{HF}$. (This method has already been used successfully in [23] to find solutions of another of the non-linear equations of atomic physics: the Thomas-Fermi [9, 33] model.)

These results for the H and HF equations, which we give in §2 were announced in [22] and sketched in [19]; seemingly unaware of our work, Bader [2] has recently presented a similar method to obtain similar results for the H (but not HF) equations. We note that prior to our work, solutions of the HF equations for a class of potentials *excluding* Coulomb potentials were found by Fonte et al. [11]. Recently, several authors [7, 8] have proved existence of the time-dependent HF equations.

In §3, we establish various “regularity” properties of any u 's (not necessarily minimizing ones), which solve the H and HF equation. Among these is the exponential falloff of the u 's announced in [22]; after our announcement, similar results *for the H equations* were obtained by [5].

In §4, we repeat the remark already made in [23] that our proof that HF theory is “exact” in the $Z \rightarrow \infty$ limit implies the same result for HF theory.

§2. Solutions of the H and HF Equations

While one could present the existence theory for the H and HF equations as two cases of one general result, we present the two theories in sequence to illustrate the extra difficulties in the HF case. The basic strategy is (cf. [23]) to introduce a weak topology on the trial functions in which the trial functions are precompact and then to prove that the functional one wishes to minimize is lower semicontinuous. This establishes that the functional is minimized at some point in the closure of the trial functions. In many cases, additional arguments are then available to prove that the minimizing point belongs to the original trial functions rather than merely to the closure.

Theorem 2.1 (H Theory). *Fix N, k ; $z_1, \dots, z_k, R_1, \dots, R_k$. There exist functions $u_1, \dots, u_N \in L^2(\mathbb{R}^3; \mathbb{C}^2)$ with $u_i \in Q(-\Delta)$, the quadratic form domain of $-\Delta$, such that the u_i minimize*

$$\begin{aligned} \tilde{\mathcal{E}}_H(u_1, \dots, u_N) = & \sum_{i=1}^N (u_i, (-\Delta + V)u_i) \\ & + \sum_{i < j} \int |u_i(x)|^2 |u_j(y)|^2 |x - y|^{-1} d^3x d^3y \end{aligned} \quad (8)$$

with the subsidiary conditions, $u_i \in Q(-\Delta)$ and

$$\|u_i\| \leq 1.$$

The u_i 's satisfy (5a) with the additional condition, that for each i , either $\varepsilon_i \leq 0$ or $u_i = 0$. In either event $\varepsilon_i = \inf \text{spec}(\tilde{h}_i)$ and if $\varepsilon_i < 0$, $\|u_i\| = 1$. If, moreover, $N < Z + 1$, then all $\varepsilon_i < 0$ and each $\|u_i\| = 1$. In all cases

$$E_H \equiv \min \{ \tilde{\mathcal{E}}_H(u_1, \dots, u_N) \mid \|u_i\| \leq 1, u_i \in Q(-\Delta) \}$$

is finite.

Remark. We have introduced the function $\tilde{\mathcal{E}}_H$ which agrees with $\mathcal{E}_H$ only when all $\|u_i\| = 1$. Theorem 2.1 says that $\tilde{\mathcal{E}}_H$ always has a minimum if we only impose $\|u_i\| \leq 1$. When the minimum of $\mathcal{E}_H$ occurs for $\|u_i\| = 1$, all i , as we assert it does if $Z + 1 > N$, then, of course, these u_i also minimize $\mathcal{E}_H$ subject to $\|u_i\| = 1$.

Proof. By a well-known result of Kato [18], for any $\varepsilon > 0$,

$$(u, Vu) \leq \varepsilon(u, -\Delta u) + C_\varepsilon(u, u)$$

from which it follows that

$$E_H = \inf \{ \tilde{\mathcal{E}}_H(u_1, \dots, u_N) \mid \|u_i\| \leq 1, u_i \in Q(-\Delta) \}$$

is finite and that for some K :

$$\tilde{\mathcal{E}}_H(u_1, \dots, u_N) \leq E_H + 1; \|u_j\| \leq 1 \Rightarrow \|Vu_j\| \leq K. \quad (9)$$

Now pick sets $u_i^{(n)}$, $1 \leq i \leq N$, $n = 1, \dots$ so that $\tilde{\mathcal{E}}_H(u_i^{(n)}) \leq E_H + 1/n$. By (9), the $u_i^{(n)}$'s lie in a fixed ball in the Sobolev space [1], $H^1 = \{u \mid \|u\| \equiv (\|u\|^2 + \|Vu\|^2)^{1/2} < \infty\}$. Thus, by the Banach-Alaoglu theorem, there exists a subsequence such that $u_i^{(n)} \rightarrow u_i^{(\infty)}$ in the weak- H^1 topology. Clearly $\|u_i^{(\infty)}\| \leq 1$. We claim that $\tilde{\mathcal{E}}_H(u_i^{(\infty)}) \leq \liminf \tilde{\mathcal{E}}_H(u_i^{(n)}) = E_H$, whence it follows that the $u_i^{(\infty)}$ minimize $\tilde{\mathcal{E}}_H$. Positive definite quadratic forms are always non-increasing under weak-limits (see e.g. [23]) so that

$$\begin{aligned} (u_i^{(\infty)}, -\Delta u_i^{(\infty)}) &\leq \liminf (u_i^{(n)}, -\Delta u_i^{(n)}) \\ (u_i^{(\infty)} u_j^{(\infty)}, |x_i - x_j|^{-1} u_i^{(\infty)} u_j^{(\infty)}) &\leq \liminf (u_i^{(n)} u_j^{(n)}, |x_i - x_j|^{-1} u_i^{(n)} u_j^{(n)}) \end{aligned}$$

since $u_i^{(n)} u_j^{(n)} \xrightarrow{w} u_i^{(\infty)} u_j^{(\infty)}$ in $L^2(\mathbb{R}^6)$. Finally, because [18, 25] V is relatively $-\Delta$ form compact [i.e. $(\Delta + 1)^{-1/2} V (\Delta + 1)^{-1/2}$ is compact] $(u_i^{(n)}, Vu_i^{(n)}) \rightarrow (u_i^{(\infty)}, Vu_i^{(\infty)})$. It follows that $\liminf \tilde{\mathcal{E}}_H(u_i^{(n)}) \geq \tilde{\mathcal{E}}_H(u_i^{(\infty)})$. Henceforth, u_i is used to denote this $u_i^{(\infty)}$.

To see that the u_i 's satisfy (5a), fix $u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N$ and let

$$\begin{aligned} f(u) &= \tilde{\mathcal{E}}_H(u_1, \dots, u_{i-1}, u, u_{i+1}, \dots, u_N) \\ &= \text{const} + (u, \tilde{h}_i u). \end{aligned}$$

Since $f(u)$ is minimized by $u = u_i$ subject to $\|u\| \leq 1$, we conclude that either $\tilde{h}_i \geq 0$, $u_i = 0$, or $\tilde{h}_i u_i = \varepsilon_i u_i$ with $\varepsilon_i = \inf \text{spec } \tilde{h}_i \leq 0$.

Now suppose that $N < Z + 1$. Let v be a spherically symmetric function on $\mathbb{R}^3$. Then $(v, \tilde{h}_i v) = (v, -\Delta v) + (v, Kv)$, where

$$K(r) = - \sum_{j=1}^k z_j (\max(r, |R_j|))^{-1} + \sum_{j \neq i} \int |u_j(k)|^2 (\max(x, r))^{-1} dx.$$

Since $\|u_j\| \leq 1$, we have that

$$K(r) \leq -[Z - (N - 1)]|r|^{-1} \quad \text{when } r > \max(|R_j|).$$

It is easy to see (use explicit hydrogenic wave functions, or a scaling argument [28]), that $(v, h_i v) < 0$ for suitable v 's. It follows that $\varepsilon_i < 0$ so that $\|u_i\| = 1$. $\square$

Remark. 1) In particular, in the neutral case $\sum z_j = N$, a solution of the H equation exists.

2) Notice that no assertion is made about uniqueness.

3) In the above proof, we used $|x_i - x_j|^{-1} \geq 0$ pointwise. In distinction, at the

analogous point in the TF theory we used the fact that $|x|^{-1}$ is positive definite.

4) The above method fails for the Hartree-like Choquard functional $\mathcal{E}(u, v) = \|Vu\|^2 + \|Vv\|^2 - \int |u(x)|^2 |v(y)|^2 |x-y|^{-1} dx dy$, because the last term is negative instead of positive. Nevertheless, alternate methods involving rearrangement inequalities can be used to prove that minimizing u and v exist, see Lieb [21].

5) Since the u_i 's are ground states of h_i , they are pointwise positive [25].

To prove the existence of solutions of the HF equation, we must extend $\mathcal{E}_{\text{HF}}$ in a manner analogous to (8); we define:

$$\tilde{\mathcal{E}}_{\text{HF}}(u_1, \dots, u_N) = \sum_{i=1}^N (u_i, (-\Delta + V)u_i) + \sum_{1 \leq i < j \leq N} ((ij)_2, |x_i - x_j|(ij)_2), \quad (10)$$

where

$$(ij)_2(x_i, x_j) = 2^{-1/2} [u_i(x_i)u_j(x_j) - u_j(x_i)u_i(x_j)].$$

For future reference we note that

$$\begin{aligned} ((ij)_2, |x_i - x_j|^{-1}(ij)_2) &= D_{ij} - E_{ij} \\ D_{ij} &= \int |x - y|^{-2} |u_i(x)|^2 |u_j(y)|^2 dx dy \\ E_{ij} &= \int |x - y|^{-1} \overline{\varrho(x)} \varrho(y) dx dy \\ \varrho(x) &= u_i(x) \overline{u_j(x)}. \end{aligned} \quad (11)$$

The critical element in the extension will be to locate the weak closure of $\{(u_1, \dots, u_N) | (u_i, u_j) = \delta_{ij}\}$:

Lemma 2.2. *Let $u_i^{(n)} \rightarrow u_i$ ($i = 1, \dots, N$) weakly with $(u_i^{(n)}, u_j^{(n)}) = \delta_{ij}$. Then $(u_i, u_j) = M_{ij}$ is an $N \times N$ matrix with $0 \leq M \leq 1$. More generally the conclusion remains true if the weaker hypothesis $(u_i^{(n)}, u_j^{(n)}) = M_{ij}^{(n)}$ with $0 \leq M^{(n)} \leq 1$ is imposed.*

Remark. The point is that it is easy to see that every $(u_1, \dots, u_N)$ with M_{ij} obeying $0 \leq M \leq 1$ arises as a weak limit of orthonormal N -tuples. Since we do not need this below, we do not give the easy proof of this converse which is based on diagonalizing M .

Proof. Let $z \in \mathbb{C}^N$. Then $(z, Mz) \equiv \sum \bar{z}_i M_{ij} z_j = (u(z), u(z)) = (w\text{-}\lim u^{(n)}(z), w\text{-}\lim u^{(n)}(z)) \leq \sum |z_i|^2$ where $u^{(n)}(z) = \sum z_i u_i^{(n)}$. The last inequality follows from the fact that balls are weakly closed and the calculation $(u^{(n)}(z), u^{(n)}(z)) = \sum |z_i|^2$. Thus $M \leq 1$. $M \geq 0$ is trivial. $\square$

We will also need the elementary observation:

Lemma 2.3. *Let $\tilde{u}_i = \sum_j a_{ij} u_j$ where $A = \{a_{ij}\}_{1 \leq i, j \leq N}$ is a unitary $N \times N$ matrix then*

$$\tilde{\mathcal{E}}_{\text{HF}}(\tilde{u}_i) = \tilde{\mathcal{E}}_{\text{HF}}(u_i).$$

Proof. Let $K_{ij} = (u_i, (-\Delta + V)u_j)$, $\tilde{K}_{ij} = (\tilde{u}_i, (-\Delta + V)\tilde{u}_j)$, $R_{i_1 i_2 j_1 j_2} = ((i_1 j_2)_2, |x - y|^{-1}(j_1 j_2)_2)$, etc. Then $\tilde{K} = A^* K A$ so $\sum \tilde{K}_{ii} = \text{Tr}(\tilde{K}) = \text{Tr}(K) = \sum K_{ii}$. Similarly, by taking traces on the antisymmetric tensor product of $\mathbb{C}^N$ with itself

$$\sum_{i < j} \tilde{R}_{ij, ij} = \sum_{i < j} R_{ij, ij}. \quad \square$$

Theorem 2.4 (HF Theory). *Fix $N, k; z_1, \dots, z_k, R_1, \dots, R_k$. There exist functions $u_1, \dots, u_N \in L^2(\mathbb{R}^3; \mathbb{C}^2)$ such that $u_i \in Q(-\Delta)$, the quadratic form domain of $-\Delta$, and such that the u_i minimize $\tilde{\mathcal{E}}_{\text{HF}}$ (given by (10)) with the subsidiary condition.*

$$M_{ij} \equiv (u_i, u_j) \text{ obeys } 0 \leq M \leq 1.$$

Moreover, the u 's obey $(u_i, u_j) = \lambda_i \delta_{ij}$ and satisfy the HF equations (7) with the additional condition that either $\varepsilon_i \leq 0$ or $u_i = 0$ for each i . $\varepsilon_1, \dots, \varepsilon_N$ are the N lowest points of the spectrum of h and if $\varepsilon_i < 0$, $\lambda_i = 1$. If moreover, $N < Z + 1$, then all $\varepsilon_i < 0$ and $(u_i, u_j) = \delta_{ij}$.

Proof. By mimicking the proof of Theorem 2.1, we find $u_i^{(\infty)}$ obeying $0 \leq M^{(\infty)} \leq 1$ which minimize $\tilde{\mathcal{E}}_{\text{HF}}$. In this proof, we use Lemma 2.2 to be sure that $0 \leq M^{(\infty)} \leq 1$ and the fact that $(ij)^{(n)} \rightarrow (ij)^{(\infty)}$ if $u_i^{(n)} \rightarrow u_i^{(\infty)}$.

Choose a unitary $N \times N$ matrix A so that $A^* M A$ is diagonal and let $u_i = \sum_j a_{ij} u_j^{(\infty)}$. By Lemma 2.3, $\{u_i\}$ minimizes $\tilde{\mathcal{E}}_{\text{HF}}$ also, and clearly $(u_i, u_j) = \lambda_i \delta_{ij}$.

Now $F(u) \equiv \tilde{\mathcal{E}}_{\text{HF}}(u_1, \dots, u_{i-1}, u, u_{i+1}, \dots, u_N) = \text{const} + (u, hu)$ so since $u = u_i$ minimizes $F(u)$ subject to $(u, u_j) = 0 (j \neq i)$, $(u, u) \leq 1$, u_i must be a linear combination of the N smallest eigenvectors of h with only eigenvalues ≤ 0 allowed. Since each u_i has this property, by further unitary change, the u_i 's can be made to obey $hu_i = \varepsilon_i u_i$.

To complete the proof we need only show that if $N - 1 < Z$, then h has N points of its spectrum below zero, i.e. $\dim P_{(-\infty, 0)}(h) \geq N$. Now write $h = h_0 - K_\Psi$ with K_Ψ given by (7d). By the positive definiteness of $|x|^{-1}$, K_Ψ is a positive operator, so we need only show that h_0 has N negative eigenvalues. This follows, as in the proof of Theorem 2.1 by considering spherically symmetric trial functions. $\square$

As already remarked in [22], the above method also yields solutions of *modified* HF equations in which we restrict the u 's to lie in certain sets with suitable properties. Because of the spin independence of the assumed Hamiltonian, H , we can obtain several solutions of the actual HF equation by taking each u_i to be a product of a space and spin function, for example, and demanding a particular symmetry of the spinor functions. We emphasize that the “true” minimum of Theorem 2.4 is *not known* to come from a set of u 's which are product functions. An additional restriction which is often made and for which our method applies, is to demand the spatial functions be real.

§ 3. Properties of Solutions

Theorem 3.1. *The solutions of the H equation (5) constructed in Theorem 2.1 obey:*

- The u_i are globally Lipschitz and lie in $D(\tilde{h}_i) = D(-\Delta)$.*
- Away from the points $r = R_j$, the u_i are infinitely differentiable.*
- Exponential falloff: For any $\alpha < |\varepsilon_i|^{1/2}$ there exists a C_α so that*

$$|u_i(x)| \leq C_\alpha \exp(-\alpha|x|).$$

Proof. $\tilde{h}_i = -\Delta + \tilde{V}_i$, with $\tilde{V}_i = V + R_\Psi^i$. Since $u_i \in L^2 \cap Q(-\Delta)$, by a Sobolev estimate [31] (for a simple discussion of this point, see [20]) it follows that $u_i^2 \in L^1 \cap L^{3/2-\varepsilon}$, so by Young's inequality, $R_\Psi^i \in L^\infty$. Thus $D(\tilde{h}_i) = D(-\Delta)$. Since $u_i \in D(\tilde{h}_i)$, we have u_i

continuous so that R_{ψ}^i is continuous. a) Now follows by a result of Kato [17] and c) by a result of Simon [29] (both specialized to the two-body case; see the original papers for other references including earlier results for the two body case). b) Follows by a bootstrap argument reminiscent of our argument in [23]: We exploit the following facts (see e.g. [24], Section IX.6):

a) Let Ω be a bounded open set. If $u \in L^2(\Omega)$; $-\Delta u = Wu$ and $W \in C^k(\Omega)$, then $D^\alpha u \in L^2$ for all multiindices α with $|\alpha| \leq k+2$.

b) If $D^\alpha u \in L^2(\Omega)$ for all α , then u is C^∞ on Ω .

The additional fact which we need is that if $u \in L^2(\mathbb{R}^3)$ and if $D^\alpha u \in L^2(\Omega)$, $|\alpha| \leq m$, then $g(x) = \int |u(y)|^2 |x-y|^{-1} dy$ is C^m on Ω . This follows by writing $|x|^{-1} = \varphi_1 + \varphi_2$ with $\varphi_2 \in C^\infty$ with support outside a ball of radius $\varepsilon/2$ and φ_1 supported in a ball of radius ε . Then $g(x) = \int |u(y)|^2 \varphi_1(x-y) dy + \int |u(y)|^2 \varphi_2(x-y) dy \equiv g_1 + g_2$. The g_2 term is C^∞ on all of $\mathbb{R}^3$. The g_1 term is easily seen to be C^m on those x such that $\{y | |x-y| < \varepsilon\} \subset \Omega$.

With a), b), and the above fact, u_j is C^∞ away from the R_j by an obvious inductive argument. $\square$

At first sight, the methods of Theorem 3.1 appear to be inapplicable to the HF case because of the non-local term. However, an elementary trick allows one to write the HF equations in local form; namely we consider the operator A on

$\bigoplus_{j=1}^N L^2(\mathbb{R}^3; \mathbb{C}^2)$ given by:

$$A_{ij} = \delta_{ij}(-\Delta + V(x) + R(x) - \varepsilon_i) + Q_{ij}(x)$$

with

$$Q_{ij}(x) = - \int |x-y|^{-1} \overline{u_j(y)} u_i(y) dy.$$

Then $A\Psi = 0$ where Ψ is the vector with $\Psi_i(x) = u_i(x)$. With this remark, the following can be proven by following the proof of Theorem 3.1:

Theorem 3.2. *The solutions of the HF equation (7) constructed in Theorem 2.4 obey:*

- a) *The u_i are globally Lipschitz and lie in $\mathcal{D}(h) = D(-\Delta)$.*
- b) *Away from the points $r = R_j$, the u_i are C^∞ .*
- c) *Let $k_0 = \min_i |\varepsilon_i|^{1/2}$. Then for any $\alpha < k_0$:*

$$|u_i(x)| \leq C_\alpha \exp(-\alpha|x|), \quad \text{all } i.$$

Remarks. 1) The common exponential rate of falloff which was obtained comes about because we have written the HF equation as a single multicomponent equation. However, it is evident that barring some miraculous cancellation, the u_i 's should have the same rate of falloff because in the HF equations $[(-\Delta + V)u_i](x)$ is a sum of terms containing all the $u_j(x)$'s in them. After making this remark in [22], we learned that it had already been made in the chemical physics literature [13].

2) By following the method of Kato [17] (or an alternative of Jensen [15]) one can give the precise singularity in the first derivatives of the u_i at the points R_j .

3) We believe that the u_i 's are real analytic away from the R_j 's.

4) In both Theorems 3.1 and 3.2, only the form of the equations and $u_i \in Q(-\Delta) \cap L^2$ is used. Any solutions satisfying this $Q(-\Delta)$ condition will obey the conclusions of the theorems.

§ 4. Connection with the Quantum Theory

We make explicit a remark of ours in [23]:

Theorem 4.1. *Let $E_N^Q(z_i, R_i)$ be given by (3) and define*

$$E_N^{\text{HF}}(z_i, R_i) \equiv \inf \{ (\Psi, H\Psi) \mid \Psi \in \mathcal{D}_{\text{phys}}; \|\Psi\| = 1; \Psi \text{ a Slater determinant} \}.$$

Let R_i, z_i be N dependent in the following manner:

- a) $z_i/N \rightarrow \lambda_i$.
- b) $R_1 = 0; R_j N^{-1/3} \rightarrow r_j$ or to ∞ .

Then

$$\lim_{N \rightarrow \infty} E_N^{\text{HF}}(z_i, R_i) / E_N^Q(z_i, R_i) = 1.$$

Proof. $E_N^Q \leq E_N^{\text{HF}} < 0$ by the variational principle so clearly the $\overline{\lim}$ is ≤ 1 . In our proof that TF theory is asymptotically correct (§ III of [23]) we constructed an explicit Slater determinant so that as $N \rightarrow \infty$, $(\Psi, H\Psi) / E_N^{\text{TF}} \rightarrow 1$ (where E^{TF} is the Thomas-Fermi energy). Since $E_N^Q / E_N^{\text{TF}} \rightarrow 1$ by [23], and $E_N^{\text{HF}} \leq (\Psi, H\Psi)$, the $\underline{\lim} \geq 1$. $\square$

Remarks. 1) As explained in [23], we expect $E_N^{\text{HF}} - E_N^Q = o(N^{5/3})$ and $E_N^Q = aN^{7/3} + bN^2 + cN^{5/3} + o(N^{5/3})$. The proof of these facts seems to us to be an important problem in understanding the bulk properties of large Z atoms and molecules. All we were able to obtain rigorously is the leading term $aN^{7/3}$. In [23], a conjecture, due to Scott, is made about the next term bN^2 (see [20] for more details).

2) We emphasize that Theorem 4.1 is only a limit theorem about total binding energy. It is physically more important to prove that HF theory gives asymptotically correct ionization energies.

References

- Adams, R.: Sobolev spaces. New York: Academic Press 1976
- Bader, P.: Méthode variationnelle pour l'équation de Hartree. E.P.F. Lausanne Thesis
- Bazley, N., Seydel, R.: Existence and bounds for critical energies of the Hartree operator. Chem. Phys. Letters **24**, 128—132 (1974)
- Behling, R., Bongers, A., Kuper, T.: Upper and lower bounds to critical values of the Hartree operator. University of Köln (preprint)
- Benci, V., Fortunato, D., Zirilli, F.: Exponential decay and regularity properties of the Hartree approximation to the bound state wavefunctions of the helium atom. J. Math. Phys. **17**, 1154—1155 (1976)
- Bethe, H., Jackiw, R.: Intermediate quantum mechanics. New York: Benjamin 1969
- Bove, A., DaPrato, G., Fano, G.: An existence proof for the Hartree-Fock time dependent problem with bounded two-body interaction. Commun. math. Phys. **37**, 183—192 (1974)
- Chadam, J.M., Glassey, R.T.: Global existence of solutions to the Cauchy problem for time-dependent Hartree equations. J. Math. Phys. **16**, 1122—1130 (1975)
- Fermi, E.: Un metodo statistico per la determinazione di alcune proprietà dell'atome. Rend. Acad. Nat. Lincei **6**, 602—607 (1927)
- Fock, V.: Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. Z. Phys. **61**, 126—148 (1930)
- Fonte, G., Mignani, R., Schiffner, G.: Solution of the Hartree-Fock equations. Commun. math. Phys. **33**, 293—304 (1973)

12. Gustafson, K., Sather, D.: Branching analysis of the Hartree equations. *Rend. di Mat.* **4**, 723—734 (1971)
13. Handy, N.C., Marron, M.T., Silverstom, H.J.: Long range behavior of Hartree-Fock orbitals. *Phys. Rev.* **180**, 45—47 (1969)
14. Hartree, D.: The wave mechanics of an atom with a non-coulomb central field. Part I. Theory and methods. *Proc. Comb. Phil. Soc.* **24**, 89—132 (1928)
15. Jensen, R.: Princeton University Senior Thesis, 1976
16. Kato, T.: Fundamental properties of Hamiltonian operator of Schrödinger type. *Trans. Am. Math. Soc.* **70**, 195—211 (1951)
17. Kato, T.: On the eigenfunctions of many particle systems in quantum mechanics. *Comm. Pure Appl. Math.* **10**, 151—177 (1957)
18. Kato, T.: *Perturbation theory for linear operators*. Berlin-Heidelberg-New York: Springer 1966
19. Lieb, E.H.: Thomas-Fermi and Hartree-Fock theory, *Proc. 1974 International Congress Mathematicians*, Vol. II, pp. 383—386
20. Lieb, E.H.: The stability of matter. *Rev. Mod. Phys.* **48**, 553—569 (1976)
21. Lieb, E.H.: Existence and uniqueness of minimizing solutions of Choquard's non-linear equation. *Stud. Appl. Math.* (in press)
22. Lieb, E.H., Simon, B.: On solutions to the Hartree Fock problem for atoms and molecules. *J. Chem. Phys.* **61**, 735—736 (1974)
23. Lieb, E.H., Simon, B.: The Thomas-Fermi theory of atoms, molecules, and solids. *Adv. Math.* **23**, 22—116 (1977)
24. Reed, M., Simon, B.: *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*. New York: Academic Press 1975
25. Reed, M., Simon, B.: *Methods of modern mathematical physics. IV. Analysis of operators*. New York: Academic Press 1977
26. Reeken, M.: General theorem on bifurcation and its application to the Hartree equation of the helium atom. *J. Math. Phys.* **11**, 2505—2512 (1970)
27. Schaefer III, H.F.: *The electronic structure of atoms and molecules*. Reading: Addison Wesley 1972
28. Simon, B.: On the infinitude vs. finiteness of the number of bound states of an N -body quantum system. *Helv. Phys. Acta* **43**, 607—630 (1970)
29. Simon, B.: Pointwise bounds on eigenfunctions and wave packets in N -body quantum systems. I. *Proc. Am. Math. Soc.* **42**, 395—401 (1974)
30. Slater, J.C.: A note on Hartree's method. *Phys. Rev.* **35**, 210—211 (1930)
31. Stein, E.: *Singular integrals and differentiability properties of functions*. Princeton: University Press 1970
32. Stuart, C.: Existence theory for the Hartree equation. *Arch. Rat. Mech. Anal.* **51**, 60—69 (1973)
33. Thomas, L.H.: The calculation of atomic fields. *Proc. Comb. Phil. Soc.* **23**, 542—548 (1927)
34. Wolkowsky, J.: Existence of solutions of the Hartree equations for N electrons. An application of the Schauder-Tychonoff theorem. *Ind. Univ. Math. Journ.* **22**, 551—558 (1972)

Communicated by J. Glimm

Received December 17, 1976

There Are No Unfilled Shells in Unrestricted Hartree-Fock Theory

Volker Bach,^{1,*} Elliott H. Lieb,^{1,2} Michael Loss,³ and Jan Philip Solovej²

¹Department of Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544

²Department of Mathematics, Fine Hall, Princeton University, Princeton, New Jersey 08544

³School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 29 July 1993)

We prove that in an exact, unrestricted Hartree-Fock calculation each energy level of the Hartree-Fock equation is either completely filled or completely empty. The only assumption needed is that the two-body interaction is—like the Coulomb interaction—repulsive; it could, however, be more complicated than a simple potential; e.g., it could have tensor forces and velocity dependence. In particular, the Hartree-Fock energy levels of atoms and molecules, often called shells, are never partially filled.

PACS numbers: 05.30.Fk, 31.10.+z, 31.20.Lr

The Hartree-Fock (HF) variational calculation provides an approximate determination of ground states and ground state energies of quantum mechanical systems such as atoms, molecules, solids, superconductors, nuclei, etc., and is widely used in physics and chemistry. Physicists often turn to it for qualitative, if not quantitative, guidance because particle correlations can sometimes be mimicked by a suitable one-body mean-field operator. Sometimes HF theory is used as the beginning of a more accurate calculational scheme, but our interest here is on qualitative features rather than numerics. The picture of quantum systems that HF theory yields is one of independent particle levels which are the eigenvalues of the HF operator; these levels are often called shells and, *a priori*, they may or may not be completely filled.

At the outset we should clarify our definition of HF theory. What we mean is the totally unrestricted theory in which one searches for the energetically very best determinantal wave function with no *a priori* assumption whatsoever on the orbitals (which are allowed to be general functions of space and spin). The orbitals are neither required to have any symmetry properties (such as having a well-defined angular momentum) nor are they required to be functions of space time functions of spin. In *restricted* HF theory, in which, for example, rotational symmetry is preserved by requiring every orbital to have a definite angular momentum, it is evident that there can be unfilled shells. It is true that the unrestricted HF ground state may break the symmetry of the Hamiltonian. There are, however, several examples where this reflects the correct physics. One such example is the Hubbard model at half filling. Indeed, in this case, the HF ground state has Néel antiferromagnetic order which would be lost if one insisted on preserving translational invariance [1].

In addition, the unrestricted HF theory yields a better estimate for the ground state energy than the restricted theory, of course, and, if conservation of symmetry is considered desirable, there is the assurance that at least one of the projections of the HF wave function onto its symmetry irreducible components (which will not be HF

functions in general) will have an equal or better energy than the unrestricted energy itself. This follows from the fact that if $\psi = \sum_j \psi_j$, with ψ_j being irreducible components, then

$$0 = \langle \psi | H | \psi \rangle - E_{\text{HF}} \langle \psi | \psi \rangle = \sum_j [\langle \psi_j | H | \psi_j \rangle - E_{\text{HF}} \langle \psi_j | \psi_j \rangle].$$

A very simple proof, given here, shows that the orbitals of the unrestricted HF ground state *fully* occupy every level of the HF operator up to the highest filled level; i.e., the degeneracy, if any, of the highest filled level of the HF operator is always exactly what is needed to accommodate the assumed number of particles. In brief, *shells are always filled in HF theory*. The idea of the proof is to assume that the level is not filled and then to use one of the remaining eigenfunctions of the HF operator to construct a new Slater determinant which has a strictly lower energy than the HF ground state. We also obtain a crude lower bound on the size of the gap, and there is no indication that this is a small number in the atomic case when it is compared to atomic energy scales.

The theorem and its proof given below obviously generalize to any system in which the two-body interaction V is repulsive, i.e., positive definite as an operator on the two-particle Hilbert space. In particular, V is allowed to be spin dependent, to contain projection operators, and to be velocity dependent. The electronic Coulomb repulsion, for example, satisfies this positivity condition. The one-body part of the Hamiltonian can be arbitrary. For convenience and because of its familiarity, we use a molecular Hamiltonian with fixed nuclei as an illustration—but only as an illustration. No symmetry is assumed to be present.

Thus, we consider a Hamiltonian

$$H = \sum_{i=1}^N \left(-\frac{\hbar^2}{2m} \Delta_i + U(\mathbf{r}_i) \right) + \frac{1}{2} \sum_{i \neq j} V(\mathbf{r}_i, \mathbf{r}_j)$$

acting on N -electron wave functions, i.e., wave functions $\Psi(\mathbf{r}_1, \sigma_1; \dots; \mathbf{r}_N, \sigma_N)$ that are antisymmetric with respect to interchanging $(\mathbf{r}_i, \sigma_i)$ with $(\mathbf{r}_j, \sigma_j)$. In the example of a molecule with K nuclear charges $Z_{j\ell}$ located at posi-

tions R_j , U and V would be given by $U(r) = -\sum_{j=1}^K Z_j \times e^2 |r - R_j|^{-1}$ and $V(r, r') = e^2 |r - r'|^{-1}$. In the general setting, V may also depend on spin coordinates as well, i.e., $V = V(r, \sigma, r', \sigma')$. Indeed, the operator V could even be allowed to be more general than a simple multiplication operator, but to keep our notation simple we shall not consider this case explicitly.

To obtain an approximate value for the ground state energy, $E_Q = \min(\Psi|H|\Psi)/(\Psi|\Psi)$, the HF calculation restricts attention to Slater determinants, i.e., wave functions of the form

$$\Phi(r_1, \sigma_1; \dots; r_N, \sigma_N) = (N!)^{-1/2} \text{Det}\{f_i(r_j, \sigma_j)\} \quad (1)$$

in which $f_1, \dots, f_N$ are orthonormal functions of space and spin: $\langle f_i | f_j \rangle = \delta_{ij}$. The approximate ground state energy is then given by the HF energy, which is defined to be

$$(h_\Phi f)(r, \sigma) = \left[-\frac{\hbar^2}{2m} \Delta + U(r) + \int_{\tau, \sigma'} \sum_{j=1}^N |\varphi_j(r', \tau)|^2 V(r, \sigma; r', \tau) d^3 r' \right] f(r, \sigma) - \sum_{\tau, \sigma'} \sum_{j=1}^N \varphi_j(r, \sigma) \int \varphi_j(r', \tau) f(r', \tau) V(r, \sigma; r', \tau) d^3 r', \quad (4)$$

and where $\varphi_1, \dots, \varphi_N$ denote the special N orthonormal functions comprising the energy minimizing Slater determinant Φ_{HF} . The eigenvalues, ε_k , of h_Φ give us some insight into the possible energy levels for binding an extra electron, but that is not our concern here.

Theorem.— Assume that V is positive definite, i.e., for every nonzero function ψ of two space-spin variables,

$$\sum_{\sigma, \sigma'} \int |\psi(r, \sigma; r', \sigma')|^2 V(r, \sigma; r', \sigma') d^3 r d^3 r' > 0.$$

Let φ be an eigenfunction of h_Φ with eigenvalue ε (i.e., $h_\Phi \varphi = \varepsilon \varphi$) that is orthogonal to the minimizing set $\varphi_1, \dots, \varphi_N$, i.e., $\langle \varphi | \varphi_k \rangle = 0$ for all $1 \leq k \leq N$. Then $\varepsilon > \varepsilon_k$ for all $1 \leq k \leq N$.

Before proving this theorem let us point out its main corollaries. First, it implies that the functions $\varphi_1, \dots, \varphi_N$ comprising Φ_{HF} occupy the N lowest energy levels of h_Φ , which is a fact that was also noted in [3] and [4]; the reader may or may not find this surprising, but we point out that there is no proof of this assertion without the assumption that $V \geq 0$. Our main point, however, is the second implication, which is more surprising: Φ_{HF} does not leave any degenerate level unfilled. There is a gap because $\varepsilon > \varepsilon_k$ for all $k = 1, 2, \dots, N$.

Proof of the theorem.— For notational convenience assume $\varepsilon_1 \leq \varepsilon_2 \leq \dots \leq \varepsilon_N$. We shall derive a contradiction to the assumption that $\varepsilon \leq \varepsilon_N$. First, we introduce some notation. Denote ε by ε_{N+1} and φ by φ_{N+1} . Further for all $1 \leq k \leq N+1$ and $1 \leq l \leq N+1$, we define

$$h_k = \langle \varphi_k | (-\hbar^2/2m) \Delta - U(r) | \varphi_k \rangle$$

$$E_{\text{HF}} = \min(\Phi|H|\Phi) \quad [\Phi \text{ has the form (1)}]. \quad (2)$$

Any minimizer, Φ_{HF} , i.e., a determinantal function satisfying $E_{\text{HF}} = \langle \Phi_{\text{HF}} | H | \Phi_{\text{HF}} \rangle$, is a HF ground state. It may not be unique. We remark that mathematical precision actually requires an “infimum” rather than “minimum” in (2) because a HF ground state may not exist. This will be the case, e.g., for a molecule with $N > 2Z + K$, i.e., a very negative ion [2]. For neutral or positively ionized atoms and molecules, however, it was proved in [3] that a HF ground state does exist and, at least in this case, the word “minimum” in (2) is justified.

If a HF ground state does exist, it necessarily obeys the HF (or self-consistent field) eigenfunction equations

$$h_\Phi \varphi_k = \varepsilon_k \varphi_k \quad (3)$$

for all $1 \leq k \leq N$, where h_Φ is the one-body operator defined by its action on an arbitrary function of one space-spin variable by

and

$$V_{k,l} = \sum_{\sigma, \sigma'} \int \frac{1}{2} |\varphi_k(r, \sigma) \varphi_l(r', \sigma') - \varphi_l(r, \sigma) \varphi_k(r', \sigma')|^2 \times V(r, \sigma; r', \sigma') d^3 r d^3 r'.$$

Notice that $V_{k,k} = 0$ and $V_{l,k} = V_{k,l} > 0$ if $k \neq l$ since V is positive definite.

Now let $\tilde{\Phi}$ be the Slater determinant built from $\varphi_1, \dots, \varphi_{N-1}, \varphi_{N+1}$, as in (1). One easily checks that

$$\begin{aligned} \langle \Phi_{\text{HF}} | H | \Phi_{\text{HF}} \rangle &= \sum_{k=1}^N h_k + \frac{1}{2} \sum_{k,l=1}^N V_{k,l}, \\ \langle \tilde{\Phi} | H | \tilde{\Phi} \rangle &= \sum_{k=1}^{N-1} h_k + \frac{1}{2} \sum_{k,l=1}^{N-1} V_{k,l} + h_{N+1} + \sum_{l=1}^{N-1} V_{l,N+1}, \end{aligned}$$

and, for $1 \leq k \leq N+1$,

$$h_k + \sum_{l=1}^N V_{k,l} = \langle \varphi_k | h_\Phi | \varphi_k \rangle = \varepsilon_k. \quad (5)$$

Notice that the term $l=k$ in the sum in (5) does not contribute since $V_{k,k} = 0$.

Now

$$\begin{aligned} \langle \tilde{\Phi} | H | \tilde{\Phi} \rangle &= \langle \Phi_{\text{HF}} | H | \Phi_{\text{HF}} \rangle + h_{N+1} - h_N \\ &\quad + \sum_{l=1}^{N-1} (V_{l,N+1} - V_{l,N}) \\ &= \langle \Phi_{\text{HF}} | H | \Phi_{\text{HF}} \rangle + \varepsilon_{N+1} - \varepsilon_N - V_{N,N+1} \\ &\leq \langle \Phi_{\text{HF}} | H | \Phi_{\text{HF}} \rangle - V_{N,N+1}. \end{aligned}$$

The last inequality uses the assumption $\varepsilon_{N+1} \leq \varepsilon_N$, but we then have a contradiction since Φ has an energy which is below the Hartree-Fock energy by the amount $V_{N,N+1} > 0$. QED.

The proof does not give a rigorous numerical estimate of the gap $\varepsilon_{N+1} - \varepsilon_N$, but it does show that the gap is at least $V_{N,N+1}$, which is usually not a tiny quantity for small systems. For such systems, even an "approximate" degeneracy is unlikely.

This work was supported by the U.S. National Science Foundation through the following grants: PHY90-19433 (V.B. and E.H.L.), DMS92-07703 (M.L.), and DMS92-03829 (J.P.S.).

Current address: FB Mathematik MA 7-2, Technische Universität Berlin, Strasse des 17. Juni 136, D-W-1000 Berlin, Germany.

- [1] V. Bach, E. H. Lieb, and J. P. Solovej, "Generalized Hartree-Fock Theory and the Hubbard Model," J. Stat. Phys. (to be published).
- [2] E. H. Lieb, Phys. Rev. A **29**, 3018 (1984).
- [3] E. H. Lieb and B. Simon, Commun. Math. Phys. **53**, 185 (1977); see also E. H. Lieb, in *Proceedings of the 1974 International Congress of Mathematicians* (Canadian Mathematical Congress, 1975), Vol. 2, p. 383.
- [4] O. A. Pankratov and P. P. Poparov, Phys. Lett. A **134**, 339 (1989).

Many-Body Atomic Potentials in Thomas–Fermi Theory

RAFAEL BENGURIA^{*,†}

Department of Physics, Princeton University, Princeton, New Jersey 08540

AND

ELLIOTT H. LIEB[†]

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08540

Received May 6, 1977

Many-body atomic potentials, ϵ , are functions of the nuclear coordinates, and are defined by differences of ground state energies, E , e.g., $\epsilon(1, 2) \equiv E(1, 2) - E(1) - E(2)$. We prove that in Thomas–Fermi theory the n -body potential always has the sign $(-1)^n$ for all coordinates. We also prove that the remainder in the expansion of the total energy E in terms of the ϵ 's, when truncated at the n -body terms, has the sign $(-1)^{n+1}$.

1. INTRODUCTION

Many-body potentials are a useful concept to describe the interaction of atoms. These quantities, which we denote by ϵ , are defined in terms of ground state energies, denoted by E , as follows:

To each positive integer j we associate a fixed nucleus of charge $z_j > 0$ located at R_j . The z_j 's can all be different. $E(j_1, \dots, j_n)$ denotes the ground state energy of an isolated molecule composed of n nuclei $j_1, \dots, j_n$, the electron–electron and nuclear–nuclear Coulomb repulsion being included. The molecule is assumed to be neutral so that the number of electrons is the sum of the z_j . Then the one-, two-, and three-body potentials are defined to be

$$\begin{aligned} \text{one body:} \quad & \epsilon(j) = E(j), \\ \text{two body:} \quad & \epsilon(j, k) = E(j, k) - [E(j) + E(k)], \\ \text{three body:} \quad & \epsilon(j, k, l) = E(j, k, l) - [E(j, k) + E(j, l) + E(k, l)] \\ & \quad + [E(j) + E(k) + E(l)], \end{aligned} \tag{1.1a}$$

* On leave from the Department of Physics, Universidad de Chile, Santiago, Chile.

† Work partially supported by U.S. National Science Foundation grant MCS 75-21684 A01.

and so forth. The inverse of (1.1a) is

$$E(j) = \epsilon(j),$$

$$E(j, k) = [\epsilon(j) + \epsilon(k)] + \epsilon(j, k), \quad (1.1b)$$

$$E(j, k, l) = [\epsilon(j) + \epsilon(k) + \epsilon(l)] + [\epsilon(j, k) + \epsilon(j, l) + \epsilon(k, l)] + \epsilon(j, k, l).$$

The generalization of (1.1) is given in (1.8) and (1.9). Note that the n -body potential $\epsilon(j_1, \dots, j_n)$ depends *only* on the charges and coordinates of nuclei $j_1, \dots, j_n$. Equation (1.1) should not be confused with a virial expansion. It will be our goal to show that in Thomas–Fermi (TF) theory, the sign of the n -body ϵ is always $(-1)^n$; moreover, if the series (1.1b) is truncated at the n -body terms (usually only the two-body terms are retained in practice) then the remainder has the sign of the first omitted terms, namely $(-1)^{n+1}$.

It is, of course, very difficult to calculate the E 's in the Schrödinger theory, especially if non-Coulombic effects are taken into account. With only pure Coulomb forces, even the two-body energy $E(1, 2)$ is difficult to estimate, except for small z . In fact the sign of the ϵ 's is unknown. One of the few things that can be reliably stated is that the two-body $\epsilon(1, 2)$ goes to $+\infty$ like $z_1 z_2 |R_1 - R_2|^{-1}$ as $R_1 \rightarrow R_2$; the other ϵ 's are bounded functions of the R_j .

It is an interesting, and perhaps useful, fact that if the Schrödinger energies are replaced by Thomas–Fermi energies, then the sign of the ϵ 's and the sign of the remainder in the expansion of E in terms of ϵ can be determined. These facts are Theorems 1 and 2 of this paper. The TF theory is an approximation to the Schrödinger theory and it is asymptotically exact [1] in the limit $z \rightarrow \infty$; therefore these theorems may not be entirely irrelevant. TF theory also plays a role in a recent proof of the stability of matter [2]. See also note added in proof.

Before proceeding we note that knowledge of the ϵ 's (or E 's) is, strictly speaking, insufficient to determine all quantities of physical interest. First, real nuclei are not static (i.e., infinitely massive) and the nuclear kinetic energy cannot be strictly separated from the electronic energy. Second, the electrons will not be in their ground state except at zero temperature. A refinement would be to use the free energy, F , in place of the ground state energy, E (either in Schrödinger or TF theory). The difficulty with this is that $F = -\infty$ unless the electrons are confined to a box (for the ground state no such confinement is necessary because the ground state neutral atoms are always bound states). Thus $\epsilon(1, 2)$, for example, would depend upon the size of the box (i.e., the density) and not merely upon nuclei 1 and 2. Third, one may wish to consider atoms that are not neutral, but then $\epsilon(1, 2, 3)$, for example, would not be defined by (1.1) until a prescription is given for apportioning the electrons among the subsystems (1, 2), (1, 3), and (2, 3). These remarks notwithstanding, the many-body potentials, as we have defined them, are useful for a wide variety of applications. The above-mentioned difficulties will not be considered further in this paper.

The TF energy for nuclei of charges $z_i > 0$ (which need not be integral) located at

R_i , $i = 1, \dots, n$ is defined as follows. In units in which $\hbar^2(8m)^{-1}(3/\pi)^{2/3} = 1$ (m = electron mass) and $|e| = 1$, one introduces the energy functional

$$\mathcal{E}(\rho) = \frac{3}{5} \int \rho(x)^{5/3} d^3x - \int V(x) \rho(x) d^3x + \frac{1}{2} \iint \rho(x) \rho(y) |x - y|^{-1} d^3x d^3y + \sum_{1 \leq i < j \leq n} z_i z_j |R_i - R_j|^{-1}. \quad (1.2)$$

Here $\rho(x) \geq 0$ is the electron density, and

$$V(x) = \sum_{j=1}^n z_j |x - R_j|^{-1}. \quad (1.3)$$

The minimum of $\mathcal{E}(\rho)$ over *all* nonnegative functions ρ occurs for a unique ρ [1]. This ρ has the property

$$\int \rho(x) d^3x = \sum_{j=1}^n z_j \equiv Z \quad (1.4)$$

and satisfies

$$\rho(x) = \phi(x)^{3/2}, \quad (1.5a)$$

$$0 \leq \phi(x) = V(x) - \int \rho(y) |x - y|^{-1} d^3y. \quad (1.5b)$$

Equation (1.4) states that a minimum energy TF molecule is always neutral. Equation (1.5) is the TF equation; its solution is unique given that $\int \rho$ and $\int \rho^{5/3}$ are finite. $\mathcal{E}(\rho)$ does have a unique minimizing ρ under the additional restriction that $\int \rho = \lambda$ and $\lambda < Z$, but this ρ does not satisfy (1.5). There is no minimizing ρ under the condition $\lambda > Z$, i.e., negative molecules do not exist in TF theory. Finally, the TF energy is given by

$$E(1, \dots, n) = \min_{\rho} \mathcal{E}(\rho) = \mathcal{E}(\bar{\rho}) \quad (1.6)$$

where $\bar{\rho}$ is the unique solution to (1.5).

The energy of an isolated TF atom is

$$E(1) = -Kz_1^{7/3} \quad (1.7)$$

where $K = 3.678$ by numerical computation.

Our theorems do not depend on the fact that the nuclei are point charges. They can be smeared out in any way provided the nuclear charge *densities* are nonnegative. This generalization requires an obvious redefinition of (1.2) and (1.3), but (1.4), (1.5), and Theorems 1, 2, 3, 4 still hold.

To state the theorems we must introduce some set theoretic notation. If $b = (b_1, b_2, \dots, b_k)$ is a finite subset of the positive integers, $E(b)$ denotes the TF energy (given by (1.2) and (1.6)) for the k nuclei of charges $z_{b_1}, \dots, z_{b_k}$ located at $R_{b_1}, \dots, R_{b_k}$. $|b| \equiv k$ is the cardinality of b . Φ denotes the empty set, and we define $E(\Phi) = 0$. $b \subset c$ means

that b is a proper subset of c , while $b \subseteq c$ means that b is a subset of c and can be c itself. $c \setminus b$ is the set of points which are in c and not in b . The generalization of (1.1) is

$$\epsilon(c) = \sum_{b \subseteq c} (-1)^{|b|+|c|} E(b) \quad (1.8)$$

from which it follows that

$$E(c) = \sum_{b \subseteq c} \epsilon(b) \quad (1.9)$$

since, if $a \subseteq c$,

$$\sum_{a \subseteq b \subseteq c} (-1)^{|b|} = (-1)^{|c|} \delta(a, c) \quad (1.10)$$

with δ being the Kronecker delta.

THEOREM 1 (Sign of the many-body potentials). *For all choices of the nuclear z 's and R 's and all sets, c ,*

$$(-1)^{|c|} \epsilon(c) \geq 0. \quad (1.11a)$$

More generally, if $b \subset c$ then

$$\tilde{E}(b, c) \equiv \sum_{b \subseteq a \subseteq c} (-1)^{|b|+|a|} E(a) \geq 0 \quad (1.11b)$$

whenever $|c \setminus b| \geq 2$ or $b = \Phi$ (this latter case is (1.11a)).

Remarks. (i) This can be extended to $(-1)^{|c|} \epsilon(c) > 0$, as in Ref. [1, Section V.2], when c is nonempty.

(ii) The special case $\epsilon(1, 2) > 0$ is Teller's theorem (cf. Ref. [1]).

THEOREM 2 (Remainder Theorem). *For all nuclear z 's and R 's, all sets c , and all $2 \leq \gamma \leq |c|$, the sign of*

$$E(c) - \sum_{\substack{b \subseteq c \\ |b| < \gamma}} \epsilon(b)$$

is $(-1)^\gamma$. In other words, if, in (1.9), we sum over all smaller than γ -body terms, then the remainder has the same sign as the first omitted terms.

THEOREM 3 (Monotonicity of the many-body potentials). *Suppose that $b \subset c$ and $|b| \geq 2$. Then for all nuclear z 's and R 's*

$$(-1)^{|b|} \epsilon(b) \geq (-1)^{|c|} \epsilon(c).$$

Remark. Again, the inequality can be shown to be strict. As an application of Theorems 1 and 3, the three-body potential satisfies

$$0 \geq \epsilon(1, 2, 3) \geq -\min[\epsilon(1, 2), \epsilon(1, 3), \epsilon(2, 3)].$$

A more general theorem which, as we shall see, implies the others, concerns the TF potential itself.

DEFINITION. Let c be a nonempty subset of the positive integers. $\phi(c, x)$ denotes the TF potential, given by (1.5b), at the point x for the corresponding collection of nuclei. If $c = \Phi$, $\phi(\Phi, x) \equiv 0$.

THEOREM 4. Suppose $b \subset c$. Then for all z 's and R 's and all x

$$\tilde{\phi}(b, c, x) \equiv \sum_{b \subseteq a \subseteq c} (-1)^{|a|+|b|} \phi(a, x) \leq 0. \quad (1.12)$$

Remark. As in Ref [1, Section V.2], strict inequality can be proved here, when c is nonempty.

Theorem 4 will be proved in the next section. The proof we give is patterned after the proof of Teller's lemma [1], but an additional combinatorial lemma is needed. All the necessary combinatorial facts are given in Section 3, and the required lemma is Lemma 13. It is important that the exponent p in (1.5a), namely $\frac{3}{2}$, satisfies $1 \leq p \leq 2$. Lemma 13 holds for $1 \leq p \leq 2$ [which would correspond to replacing $\int \rho^{5/3}$ by $\int \rho^k$, $\frac{3}{2} \leq k \leq 2$ in $\mathcal{E}(\rho)$], but Lemma 13 is false for $p < 1$ or $p > 2$. It is amusing that $p = 2$ corresponds to $k = \frac{3}{2}$, and TF theory does not exist for $k \leq \frac{3}{2}$ (because $\mathcal{E}(\rho)$ is then not bounded below).

We conclude this section with the mention of a basic fact [1] that relates the TF potential to the TF energy. Theorem 4 will follow from Lemmas 5 and 13.

LEMMA 5. Let c be nonempty and $j \in c$. Then the derivative of $E(c)$ with respect to the nuclear charge z_j is given by

$$\partial E(c)/\partial z_j = \lim_{x \rightarrow R_j} \{ \phi(c, x) - z_j |x - R_j|^{-1} \}.$$

This lemma is proved [1] by differentiating (1.2) with respect to z_j and using the fact that $\mathcal{E}(\rho)$ is stationary with respect to the minimizing ρ .

2. PROOFS OF THEOREMS 1, 2, 3, AND 4

In this section we give the proof of the theorems stated in the Introduction. We begin with the general Theorem 4 about the TF potential. The other theorems are a consequence of this one. The proof of Theorem 4 is based on the proof of Teller's lemma (see Theorem V.5, Ref. [1]) and the combinatorial Lemma 13 proved in Section 3. Logically, Section 3 should be inserted at this point, but the only things needed are (i) the definition of the transform (3.1); (ii) the definition of an anticanonical function, (iii) Lemma 13.

Proof of Theorem 4. Insert (1.5b) into the right side of (1.12). If $j \in c$, it is easy to check, using (1.10), that all terms of the type $z_j |x - R_j|^{-1}$ either cancel (if $c \neq$

$b \cup \{j\}$) or have a coefficient -1 (if $c = b \cup \{j\}$). The (distributional) Laplacian of $\tilde{\phi}$ therefore satisfies

$$\begin{aligned} (4\pi)^{-1} \Delta \tilde{\phi}(b, c, x) &\geq \sum_{b \subseteq a \subseteq c} (-1)^{|a|+|b|} \phi^{3/2}(a, x) \\ &= \widehat{\phi^{3/2}}(b, c, x). \end{aligned} \quad (2.1)$$

Let

$$D(b, c) = \{x \mid \tilde{\phi}(b, c, x) > 0\}.$$

Suppose we can show that

$$x \in D(b, c) \quad \text{implies} \quad \widehat{\phi^{3/2}}(b, c, x) \geq 0. \quad (2.2)$$

Then $\tilde{\phi}(b, c, \cdot)$ is subharmonic on $D(b, c)$ and therefore takes its maximum on the boundary of $D(b, c)$ or at infinity. If $|c \setminus b| > 1$, $\tilde{\phi}$ is continuous on all of $\mathbb{R}^3$; if $c = b \cup \{j\}$, $\tilde{\phi}$ is continuous on $\mathbb{R}^3 \setminus R_j$ and $\tilde{\phi}(b, c, x) \rightarrow -\infty$ as $x \rightarrow R_j$ [1]. Hence $\tilde{\phi}$ is zero on the boundary of D . Since $\tilde{\phi}$ goes to zero as $x \rightarrow \infty$ [1], we conclude that D is empty. This potential theoretic argument is the main idea in the proof of Teller's lemma [1]. Thus, the problem is reduced to proving assertion (2.2). To do so we use induction on the cardinality of $c \setminus b$. Let $P(n)$, for $n \geq 1$, be the proposition: Theorem 4 is true when $1 \leq |c \setminus b| \leq n$, i.e., $\tilde{\phi}(b, c, x) \leq 0$ for all x . If $n = 1$ then $\tilde{\phi}(b, c, x) = -\phi(c, x) + \phi(b, x)$, and (2.2) is trivially verified. Hence $P(1)$ is true. (Note: This case is just Teller's original lemma.) $P(n+1)$ will be proved by showing that $P(n)$ implies (2.2). $D(b, c)$ is empty when $b \neq \Phi$. For $x \in D(\Phi, c)$ fixed, consider $d \rightarrow \phi(d, x)$ as a positive function on the power set 2^c . In the terminology of Section 3, $\phi(\cdot, x)$ is anticanonical. Then Lemma 13 is precisely (2.2). ■

Proof of Theorem 1 (sign of the many-body potentials). Let $j \in c \setminus b$, and consider the dependence of $\tilde{E}(b, c)$ on z_j . When $z_j = 0$, $E(a) = E(a \setminus \{j\})$ if $j \in a$, and it is then easy to check that

$$\tilde{E}(b, c; z_j = 0) = 0. \quad (2.3)$$

By Lemma 5

$$\partial \tilde{E}(b, c) / \partial z_j = \lim_{x \rightarrow R_j} \sum_{b \cup \{j\} \subseteq a \subseteq c} (-1)^{|b|+|a|} \{\phi(a, x) - z_j |x - R_j|^{-1}\}. \quad (2.4)$$

Using (1.10) one sees that the terms on the right side of (2.4) proportional to $z_j |x - R_j|^{-1}$ either cancel (if $|c \setminus b| \geq 2$) or else have a coefficient $+1$ (if $c = b \cup \{j\}$). The terms involving $\phi(a, x)$ are, by definition, $-\tilde{\phi}(b \cup \{j\}, c, x)$. If $|c \setminus b| \geq 2$, the right side of (2.4) is then $-\lim_{x \rightarrow R_j} \tilde{\phi}(b \cup \{j\}, c, x)$ and this is nonnegative by Theorem 4. If $b = \Phi$ and $c = \{j\}$, the right side of (2.4) is $\lim_{x \rightarrow R_j} \{z_j |x - R_j|^{-1} - \phi(\{j\}, x)\}$ and this is positive by (1.5b). In either case, therefore,

$$\partial \tilde{E}(b, c) / \partial z_j \geq 0. \quad (2.5)$$

Equations (2.3) and (2.5) prove the theorem. ■

Proof of Theorem 3 (monotonicity of the many-body potentials). This is really a simple corollary of Theorem 1. Obviously it is sufficient to consider the case $c = b \cup \{j\}$ and $j \notin b$. Let $a = \{j\}$, whence $|c \setminus a| = |b| \geq 2$. By Theorem 1, $\tilde{E}(a, c) \geq 0$. Inserting (1.9) into the definition of $\tilde{E}(a, c)$, and interchanging the summation order, we have that

$$\begin{aligned}
 0 \leq \tilde{E}(a, c) &= \sum_{h \subseteq c} \left(\sum_{a \cup h \subseteq d \subseteq c} (-1)^{|d|} \right) (-1)^{|a|} \epsilon(h) \\
 &= \sum_{h \subseteq c} \delta(a \cup h, c) (-1)^{|c|+|a|} \epsilon(h) \\
 &= \sum_{c \setminus a \subseteq h \subseteq c} (-1)^{|c|+|a|} \epsilon(h) \\
 &= (-1)^{|b|} \epsilon(b) - (-1)^{|c|} \epsilon(c). \quad \blacksquare
 \end{aligned} \tag{2.6}$$

Proof of Theorem 2 (remainder theorem). For $b \subseteq c$ and $\gamma \geq 0$ introduce

$$l_\gamma(b, c) = (-1)^{\gamma+|b|} \sum_{\substack{b \subseteq f \subseteq c \\ |f \setminus b| \geq \gamma}} \epsilon(f). \tag{2.7}$$

We want to show that $l_\gamma(\Phi, c) \geq 0$ for $\gamma \geq 2$. First consider l_0 . In general,

$$l_0(b, c) = \tilde{E}(c \setminus b, c) \tag{2.8}$$

by the calculation in (2.6) (which holds for all $a \subseteq c$). Since $c \setminus (c \setminus b) = b$, we conclude (from Theorem 1) that $l_0(b, c) \geq 0$ if either $c = b$ or $|b| \geq 2$. Now it is easy to check that for $\gamma \geq 1$ and $j \in c \setminus b$:

$$l_\gamma(b, c) = l_\gamma(b, c \setminus \{j\}) + l_{\gamma-1}(b \cup \{j\}, c). \tag{2.9}$$

When $\gamma \geq 1$, $l_\gamma(c, c) = 0$ (all c) by (2.7). If $\gamma = 1$ and $b \neq \Phi$, a simple induction on $|c \setminus b|$ using (2.9) and the fact that $|b \cup \{j\}| \geq 2$, shows that $l_1(b, c) \geq 0$. Now suppose $\gamma \geq 2$ (all b, c). Using (2.8) and induction on γ followed by induction on $|c \setminus b|$ together with the fact that $b \cup \{j\} \neq \Phi$ in (2.9), one has that $l_\gamma(b, c) \geq 0$ for all b, c when $\gamma \geq 2$. $\blacksquare$

3. FUNCTIONS DEFINED ON THE POWER SET OF A GIVEN SET

In this section we present some general properties of positive real functions defined on the power set of a given set.

DEFINITION. Consider a fixed, nonempty, finite set S and its power set 2^S , that is the set of all subsets of S (including the empty set Φ). To every function f from 2^S to

the complex numbers $\mathbb{C}$ we associate the function (“transform of f ”) $\tilde{f}: 2^S \times 2^S \rightarrow \mathbb{C}$, defined by

$$\tilde{f}(a, b) = \sum_{a \subseteq d \subseteq b} (-1)^{|a|+|d|} f(d) \quad (3.1)$$

when $a \subseteq b \subseteq S$. If $a \not\subseteq b$ $\tilde{f}$ is defined to be zero. Note that $\tilde{f}(a, a) = f(a)$.

The following lemma provides a convolution formula for the “transforms”

LEMMA 6. *Let f, g be functions on 2^S . Then*

$$\widetilde{f \cdot g}(a, b) = \sum_{a \subseteq d \subseteq b} \tilde{f}(a, d) \tilde{g}(d, b). \quad (3.2)$$

Remark. Although the left side of (3.2) is invariant under the interchange of f and g , the right side is not manifestly invariant.

Proof. Inserting definition (3.1) into the RHS of (3.2) and interchanging the sum orders we get

$$\begin{aligned} \text{RHS (3.2)} &= \sum_{a \subseteq d \subseteq b} \left(\sum_{a \subseteq e \subseteq d} (-1)^{|a|+|e|} f(e) \right) \left(\sum_{d \subseteq h \subseteq b} (-1)^{|d|+|h|} g(h) \right) \\ &= \sum_{a \subseteq e \subseteq b} \sum_{e \subseteq h \subseteq b} \left(\sum_{e \subseteq d \subseteq h} (-1)^{|d|} \right) (-1)^{|a|+|e|+|h|} f(e) g(h) \\ &= \sum_{a \subseteq e \subseteq b} \sum_{e \subseteq h \subseteq b} (-1)^{|a|+|e|} \delta(e, h) f(e) g(h) \\ &= \sum_{a \subseteq e \subseteq b} (-1)^{|a|+|e|} f(e) g(e) = \widetilde{f \cdot g}(a, b). \end{aligned}$$

In the third equality we have used Eq. (1.10).

DEFINITION. $P(S)$ denotes the nonnegative functions $f: 2^S \rightarrow \mathbb{R}^+$. If $f \in P(S)$, note that $\tilde{f}(a, a) \geq 0$. We call $f \in P(S)$ *normal* if $\tilde{f}(a, b) \leq 0$ for every $a \subset b \subseteq S$ such that $b \setminus a \neq S$. f is *canonic* (resp. *anticanonic*) if f is normal and $\tilde{f}(\Phi, S) \leq 0$ (resp. if $\tilde{f}(\Phi, S) \geq 0$).

We will now prove some properties of these special functions.

LEMMA 7. *Suppose that $f \in P(S)$ and either (i) f is canonic and $|S| \geq 1$ or (ii) f is normal and $|S| \geq 2$. Then $a \subset b \subseteq S$ implies that $\tilde{f}(a, b) \leq 0$.*

Proof. It is sufficient to consider $a \subset S$ and $b = a \cup \{x\}$ with $x \in S, x \notin a$. Then, in both cases,

$$0 \geq \tilde{f}(a, b) = f(a) - f(b). \quad \blacksquare$$

LEMMA 8. *If $f \in P(S)$ is normal (resp. canonic) then $f^{1/2}$ is normal (resp. canonic).*

Proof. By hypothesis, $\tilde{f}(a, b) \leq 0$ whenever $a \subset b \subseteq S$ and $|b \setminus a| < |S|$. We want to show that under the same conditions $\tilde{f}^{1/2}(a, b) \leq 0$. The proof is by induction on the cardinality of $|b \setminus a|$. Consider the proposition $P(n)$ for $n \leq |S|$:

$$\tilde{f}^{1/2}(a, b) \leq 0 \quad \text{for } a \subset b \subseteq S, \quad |b \setminus a| \leq n.$$

$P(1)$ is true by inspection. We will show $P(n)$ implies $P(n+1)$ when $(n+1) < |S|$. The convolution formula (3.2) can be written as

$$\tilde{f}(a, b) = \sum_{a \subset k \subset b} \tilde{f}^{1/2}(a, k) \tilde{f}^{1/2}(k, b) + \tilde{f}^{1/2}(a, b) [\tilde{f}^{1/2}(a, a) + \tilde{f}^{1/2}(b, b)]. \quad (3.3)$$

Assume $|b \setminus a| = n+1$. The sum appearing in Eq. (3.3) is nonnegative by $P(n)$. Moreover $\tilde{f}(a, b) \leq 0$ (if $n+1 < |S|$). Therefore $\tilde{f}^{1/2}(a, b) \leq 0$ if $f(a) + f(b) > 0$; otherwise let $f(c) \rightarrow f(c) + x$ and use continuity in x . Finally, (3.3) implies $P(|S|)$ if f is canonic. ■

The following is an immediate consequence of Lemma 8.

COROLLARY 9. *If f is normal (resp. canonic), then f^p with $p = 2^{-k}$, k a positive integer, is normal (resp. canonic).*

Proof. By Lemma 8 and induction on k . ■

Remarks. (i) Note that if f is anticanonic the only thing we can say about $f^{1/2}$ is that it is normal.

(ii) Our goal is to extend Lemma 8 to f^p , $p \in [0, 1]$ and, indeed to any positive Pick function. This is Lemma 12. Lemma 8 is not needed for the proof of Lemma 12, but we presented it for two reasons: (a) the case $p = \frac{1}{2}$ is what is needed for TF theory; (b) the proof just given for $p = \frac{1}{2}$ (and hence $p = 2^{-k}$) is simpler than the proof of Lemma 12.

LEMMA 10. *If f is normal and $f(a) > 0$, all $a \subseteq S$ then $\tilde{f}^{-1}(a, b) > 0$ for $a \subseteq b \subseteq S$, and $b \setminus a \neq S$. If f is also canonic, then $\tilde{f}^{-1}(a, b) \geq 0$ for $a \subseteq b \subseteq S$.*

Proof. Again we use induction on the cardinality of $b \setminus a$ and the convolution Eq. (3.2).

$$\delta(a, b) = \tilde{1}(a, b) = \sum_{a \subset d \subset b} \tilde{f}(a, d) \tilde{f}^{-1}(d, b) + \tilde{f}(a, a) \tilde{f}^{-1}(a, b) + \tilde{f}(a, b) \tilde{f}^{-1}(b, b).$$

The proof is now a straightforward imitation of the proof of Lemma 8. ■

Consider $f \in P(S)$ and its transform $\tilde{f}$. If $x \in P(S)$ is a constant mapping (i.e., $x(a) = x \geq 0$, all $a \subseteq S$) then

$$(\widetilde{f+x})(a, b) = \tilde{f}(a, b), \quad a \subset b \subseteq S \quad (3.4)$$

as can be easily checked using (1.10) and the definition of the transform (3.1). Therefore if f is respectively canonic, anticanonic, or normal, so is $f + x$.

LEMMA 11. *Let $f \in P(S)$ be normal (resp. canonic). Let $x > 0$. Then the function g given by $g(a) = f(a)[f(a) + x]^{-1}$ is normal (resp. canonic).*

Proof. We have to show that $\tilde{g}(a, b) \leq 0$ for $a \subset b \subseteq S$ (with the restriction $b \setminus a \neq S$ if f is only normal).

$$\tilde{g}(a, b) = \tilde{f}(a, b) - \widetilde{x(f+x)^{-1}}(a, b).$$

The lemma is proved by using (3.4), Lemma 10, and $\tilde{f}(a, b) = \delta(a, b)$. ■

Finally we state the generalization of Lemma 8.

LEMMA 12. *If f is canonic (resp. normal), then f^p is canonic (resp. normal) for $0 \leq p \leq 1$.*

Proof. If $p = 0$ or 1 the proof is trivial. For $0 < p < 1$, use the representation

$$f^p = K_p \int_0^\infty dx x^{p-1} f(f+x)^{-1} \quad (3.5)$$

(with $K_p = \pi^{-1} \sin(p\pi)$, $0 < p < 1$). Taking the transform on both sides of (3.6) the lemma follows from Lemma 11. ■

Remark. Lemma 12 obviously remains true if f^p is replaced by $h(f)$, where $h: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ has the representation

$$h(x) = a + bx + \int_0^\infty x(x+y)^{-1} d\mu(y)$$

with $a, b \geq 0$ and μ a positive Borel measure on $[0, \infty]$. Such functions are Pick (or Herglotz) functions [3] on $\mathbb{R}^+$.

With the help of the previous lemmas we can now prove the following theorem which is needed in the previous section.

LEMMA 13. *If f is anticanonic and $0 \leq p \leq 1$ then $\widetilde{f^{1+p}}(\Phi, S) \geq 0$.*

Proof. If $|S| = 1$ the proof is by inspection. If $|S| \geq 2$ let us define g by

$$f(a) = g(a) + f(\Phi).$$

Then $g \in P(S)$, by Lemma 7 and, moreover, $g(\Phi) = 0$. We call $f(\Phi) = x_0$ ($x_0 \geq 0$). Define the function $g_x = g + x$ on 2^S . By (3.4), g_x is anticanonic for all $x \geq 0$. Consider $h: \mathbb{R}^+ \rightarrow \mathbb{R}$ defined by

$$h(x) = \widetilde{g_x^{1+p}}(\Phi, S). \quad (3.6)$$

Since $f = g_x$ for $x = x_0$, we want to show that $h(x) \geq 0$ for every nonnegative x . From (3.6)

$$h'(x) = (1 + p) \widetilde{g}_x^p(\Phi, S). \quad (3.7)$$

Again using the convolution Eq. (3.2) we have that

$$h(x) = \sum_{\Phi \subset a \subset S} \widetilde{g}_x(\Phi, a) \widetilde{g}_x^p(a, S) + (1 + p)^{-1} h'(x) \widetilde{g}_x(\Phi, \Phi) + \widetilde{g}_x(\Phi, S) \widetilde{g}_x^p(S, S). \quad (3.8)$$

Note that $\widetilde{g}_x(\Phi, \Phi)$ and $\widetilde{g}_x(S, S)$ are nonnegative. Furthermore, the sum in (3.8) is nonnegative by Lemma 12. If $p = 1$ then $(1 + p)^{-1} h'(x) = \widetilde{g}_x(\Phi, S)$ and the two last terms in (3.8) are $\widetilde{g}_x(\Phi, S)[\widetilde{g}_x(\Phi, \Phi) + \widetilde{g}_x(S, S)] \geq 0$ for all x . This proves the lemma when $p = 1$, and henceforth we assume $p < 1$. Now

$$h(0) = \sum_{\Phi \subset a \subset S} \widetilde{g}(\Phi, a) \widetilde{g}^p(a, S) + \widetilde{g}(\Phi, S) \widetilde{g}^p(S, S)$$

which is positive because g is anticanonic (and therefore normal). As $x \rightarrow \infty$, $g_x^{1+p} \sim x^{1+p} + (1 + p) x^p g + o(1)$. Hence, for large x ,

$$h(x) \sim (1 + p) x^p \widetilde{g}(\Phi, S) \geq 0$$

because g is anticanonic. $h(x)$ is a continuously differentiable function of x on $(0, \infty)$, $h(0) \geq 0$ and, as $x \rightarrow \infty$, either $h(x) \rightarrow 0$ or $h(x) \rightarrow +\infty$. Therefore, either $h(x) \geq 0$ for every positive x or there must exist $y > 0$ such that $h(y) < 0$ and $h'(y) = 0$. But $h'(y) = 0$ implies $h(y) > 0$ (by Eq. (3.8)). ■

Remarks. (i) If p is outside the interval $[0, 1]$ and $|S| \geq 3$ the statement given in Lemma 13 is definitely false. Consider the example

$$\begin{aligned} S &= \{1, 2, 3\} \\ f(1, 2, 3) &= 1, & f(\Phi) &= 0, \\ f(1, 2) &= f(1, 3) = \frac{3}{4}, & f(2, 3) &= \frac{1}{2}, \\ f(1) &= \frac{1}{2}, & f(2) = f(3) &= \frac{1}{4}. \end{aligned}$$

Here f is anticanonic but $\widetilde{f}^{2+\epsilon}(\Phi, S) < 0$ for every $\epsilon > 0$.

(ii) If $|S| = 2$, a simple convexity argument shows that the statement made in Lemma 13 is valid for any $p > 0$. In the case $S = \{1, 2\}$ f anticanonic is equivalent to $f(1, 2) \geq f(1) \geq f(\Phi)$; $f(1, 2) \geq f(2) \geq f(\Phi)$; $f(1, 2) + f(\Phi) \geq f(1) + f(2)$. We want to show that $f(1, 2)^q + f(\Phi)^q \geq f(1)^q + f(2)^q$, when $q = 1 + p \geq 1$. Now $x \rightarrow x^q$ is convex and monotone increasing for $x \geq 0$. For any such function, $g, w \leq x \leq y \leq z$, and $w + z \geq x + y$ imply $g(w) + g(z) \geq g(x) + g(y)$.

Note added in proof. The dependence of E on the R_i , for fixed z_i , is also an interesting question in TF theory. We have been able to prove [4] that the pressure and compressibility (defined by the change of E under uniform dilation) are positive, and that the kinetic energy is superadditive. This is true even for finite molecules. Thus, problems 6, 7, and 8 of Ref. [1, p. 33] and the problems posed in Ref. [1, pp. 104–105] have been solved affirmatively.

REFERENCES

1. E. H. LIEB AND B. SIMON, *Advances in Math.* **23** (1977), 22–116. See also, E. H. LIEB AND B. SIMON, *Phys. Rev. Lett.* **31** (1973), 681–683; E. H. LIEB, “Proc. Int. Congress of Math.,” Vancouver (1974); *Rev. Modern Phys.* **48** (1976), 553–569.
2. E. H. LIEB AND W. E. THIRRING, *Phys. Rev. Lett.* **35** (1975), 687–689. See also, E. H. LIEB, *Rev. Modern Phys.* **48** (1976), 553–569.
3. W. F. DONOGHUE, JR., “Monotone Matrix Functions and Analytic Continuation,” Springer, New York, 1974.
4. R. BENGURIA AND E. H. LIEB, The positivity of the pressure in Thomas–Fermi theory, *Commun. Math. Phys.*, to be submitted.

The Positivity of the Pressure in Thomas Fermi Theory*

R. Benguria^{1**} and E. H. Lieb²

¹ Department of Physics and

² Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08 540, USA

Abstract. We prove the positivity of the pressure and compressibility for neutral systems in the Thomas-Fermi theory of molecules. Our results include some new properties of the Thomas-Fermi potential and a proof that the kinetic energy is superadditive.

I. Introduction

The Thomas-Fermi (TF) theory of atoms, molecules and solids has been given a firm mathematical foundation and many of the qualitative properties of the theory are understood and have been proven [1] (see also [2]; properties of the many-body TF potential are proved in [3]). There were, however, some open questions in [1], one of which we solve in this paper: the positivity of the pressure and compressibility for neutral systems.

The TF theory is defined by the energy functional (in units in which $h^2(8m)^{-1}(3/\pi)^{2/3} = 1$ and $|e| = 1$, where e and m are the electron charge and mass)

$$\xi(\varrho) = K(\varrho) - A(\varrho) + R(\varrho) + U \quad (1.1)$$

$$K(\varrho) = \frac{3}{5} \int \varrho(x)^{5/3} dx$$

$$A(\varrho) = \int V(x)\varrho(x)dx$$

$$V(x) = \sum_{j=1}^k z_j |x - R_j|^{-1}$$

$$R(\varrho) = \frac{1}{2} \iint \varrho(x)\varrho(y) |x - y|^{-1} dx dy$$

$$U = \sum_{1 \leq i < j \leq k} z_i z_j |R_i - R_j|^{-1}. \quad (1.2)$$

Here $z_1, \dots, z_k \geq 0$ are the charges of k fixed nuclei located at $R_1, \dots, R_k$. $\int dx$ is always a three-dimensional integral. $\xi(\varrho)$ is defined for electron densities $\varrho(x) \geq 0$ such that $\int \varrho$ and $\int \varrho^{5/3}$ are finite.

* Work partially supported by U.S. National Science Foundation grant MCS 75-21684 A02

** On leave from Department of Physics, Universidad de Chile, Santiago, Chile

The TF energy for λ (not necessarily integral) electrons is defined by

$$e(\lambda) = \inf \{ \xi(\varrho) \mid \int \varrho = \lambda \} \quad (1.3)$$

It is known [1] that for $\lambda \leq Z \equiv \sum_{j=1}^k z_j$ there is a unique minimizing ϱ for (1.3). It is the unique solution to the TF equation

$$\varrho(x)^{2/3} = \max[\phi(x) - \Phi_0, 0] \quad (1.4a)$$

for some $\Phi_0 \geq 0$, and with

$$\phi(x) \equiv V(x) - \int |x - y|^{-1} \varrho(y) dy. \quad (1.4b)$$

$-\Phi_0$ is the chemical potential [1], i.e.

$$\frac{de(\lambda)}{d\lambda} = -\Phi_0. \quad (1.5)$$

For $\lambda \leq Z$, $\phi(x) > 0$, all x . $\Phi_0 = 0$ if and only if $\lambda = Z$ and hence, for the *neutral case* the TF equation is

$$\varrho^{2/3}(x) = \phi(x). \quad (1.4c)$$

If $\lambda > Z$, there is no minimizing ϱ for (1.3), and $e(\lambda) = e(Z)$ in this case.

There are various possible definitions of the pressure. The one we shall use is the “*change in energy under uniform dilation*” defined as follows: Replace each R_i by lR_i , l being a scale factor, and let $e(\lambda, l)$ be the TF energy for a given λ and l . Then $P = -\partial e / \partial V$ which we interpret as

$$P = -(3l^2)^{-1} \partial e(\lambda, l) / \partial l. \quad (1.6)$$

The reciprocal compressibility, κ^{-1} , should be $-V \frac{\partial P}{\partial V}$ which we interpret as

$$\kappa^{-1} = -(l/3) \partial P / \partial l. \quad (1.7)$$

We shall prove that in the neutral case P and κ^{-1} are nonnegative (in the atomic case they are, of course, zero). In the process of doing so, we shall prove several interesting facts about the dependence of $\phi(x)$, K , A and R on the z_i . (Note: here and in the sequel, $\phi(x)$, K , A , R , etc. mean the respective quantities evaluated at the unique, minimizing TF density, ϱ .) We are not able to prove that P and κ are non-negative in the ionic (i.e. subneutral) case but *conjecture* that they are. The only thing we shall have to say about the ionic case except for appendix B is to give a formula (1.14) for P in terms of e and K . We are led to make the further *conjecture* that P is a decreasing function of λ and thus that the neutral case is the worst case. When $\lambda = 0$, $P > 0$ and $\kappa > 0$ because $e = l^{-1} \sum_{i < j} z_i z_j |R_i - R_j|^{-1}$. In other words, the pressure is positive because the nuclear repulsion dominates the attractive forces; this repulsion presumably grows stronger as electrons are removed from the system.

The above definitions (1.6, 1.7) of P and κ carry over, in the thermodynamic limit, to the ordinary definitions for a solid (see [1], Sect. VI). There are, however,

two other useful definitions which we will not touch upon in the main text (see Appendix B, however) except to *conjecture* that P and κ are non-negative for these definitions as well.

(i) *Dilation in one direction*: Let $R_i \rightarrow (lR_i^1, R_i^2, R_i^3)$, instead of $R_i \rightarrow lR_i$. Since this is a one dimensional expansion it seems appropriate to define $P = -\partial e(\lambda, l)/\partial l$ and $\kappa^{-1} = -l\partial P/\partial l$. As far as nonnegativity is concerned, this new definition changes κ but not P .

(ii) *Separation relative to a plane*: choose any plane which does not contain nuclei. For convenience it may be assumed to be the x – y plane $\{(x^1, x^2, x^3)|x^3=0\}$. If $R_i = (R_i^1, R_i^2, R_i^3)$, replace R_i^3 by $R_i^3 + l$ if $R_i^3 > 0$ and by $R_i^3 - l$ if $R_i^3 < 0$. Note that in this case we shift by l instead of dilate by l . Again, $P = -\partial e(\lambda, l)/\partial l$ and $\kappa^{-1} = -l\partial P/\partial l$. In appendix B we will prove that $P > 0$ if the plane is a symmetry plane. This latter case was also proved by Balász [4] but our proof is somewhat different; it uses reflection positivity. Balász assumed there were only two nuclei, but his method works for any symmetric situation. One reason for being interested in this special case is that our (and Balász') proofs are valid for the *ionic* case as well.

The definitions we shall work with (1.6, 1.7) have one virtue, namely the dependence of e on l can be converted into a dependence of e on the z_i . This is a consequence of the following *scaling properties*:

Henceforth, $R_1, R_2, \dots, R_k$ are fixed (with $R_i \neq R_j$ if $i \neq j$). We denote the k -tuple $z_1, \dots, z_k$ simply by $\underline{z}$. Let $e(\underline{z}, \lambda, l)$ be the energy with the uniform dilation l . Then, from (1.1),

$$e(\underline{z}, \lambda, l) = l^{-7} e(l^3 \underline{z}, l^3 \lambda, 1) \quad (1.8)$$

and the minimizing TF density satisfies

$$\varrho(\underline{z}, \lambda, l; x) = l^{-6} \varrho(l^3 \underline{z}, l^3 \lambda, 1; x/l). \quad (1.9)$$

Substituting (1.8) in (1.6, 1.7) yields (assuming that all derivatives exist)

$$3l^{10}P = 7e - 3l^3 \sum_{i=1}^k z_i e_i - 3l^3 \lambda e_2 \quad (1.10)$$

$$\begin{aligned} 9l^{10}\kappa^{-1} = & 70e - 42l^3 \sum_{i=1}^k z_i e_i - 42l^3 \lambda e_2 \\ & + 9l^6 \sum_{i,j=1}^k z_i z_j e_{ij} + 9l^6 \lambda^2 e_{22} + 18l^6 \lambda \sum_{i=1}^k z_i e_{2i}. \end{aligned} \quad (1.11)$$

In (1.10, 1.11) the notation is the following:

$$e_i \equiv \partial e(\underline{x}, y, 1)/\partial x_i, \quad e_2 \equiv \partial e(\underline{x}, y, 1)/\partial y, \quad \text{etc.}$$

These quantities are evaluated at $\underline{x} = l^3 \underline{z}$ and $y = l^3 \lambda$. A numerical error in the expression for κ^{-1} was made in Ref. [1], Eq. (145).

A more convenient form for P is obtained by noting that $e = K - A + R + U$. Furthermore, if (1.4) is multiplied by $\varrho(x)$ and integrated over the set on which $\varrho(x) \geq 0$, one obtains

$$(5/3)K = A - 2R - l^3 \lambda \Phi_0 \quad (1.12)$$

moreover, $e_2 = -\Phi_0$ (cf. (1.5)). Finally,

$$l^3 \sum_{i=1}^k z_i e_i = 2U - A = 2e - (1/3)K + l^3 \lambda \Phi_0 \quad (1.13)$$

([1], Theorem II.16 or Lemma V.7). Combining these facts and then using (1.8), yields, for all λ ,

$$3l^3 P(z, \lambda, l) = e(z, \lambda, l) + K(z, \lambda, l). \quad (1.14)$$

For an atom, $2K = A - R$ (Virial Theorem, [1], Theorem II.22), $U = 0$ and $e = K - A + R$. Thus (1.14) gives $P = 0$ for all λ , as it should in this case.

The conjecture stated above, that the neutral case is the worst can be given a more transparent form:

Conjecture 1. $e + K$ is a decreasing function of λ for fixed z and R_i .

In this paper we will prove the positivity of P and κ for the neutral case. In the next Section the list of theorems to be proved is given. These theorems have an easy heuristic proof and these are given in Sect. III. We do so because these heuristic proofs are a guide to the proper proofs given in Sect. IV, and because they may be a useful guide to future work.

II. Theorems to Be Proved

We will be concerned only with the neutral case and use the notation $\phi(z, x)$, $\varrho(z, x)$, $e(z)$, $K(z)$, $A(z)$, $R(z)$, $U(z)$ to denote the TF potential and density at the point $x \in \mathbb{R}^3$, the total TF energy, the kinetic energy, the attractive energy, the electron repulsion and the nuclear repulsion, respectively, (cf. (1.1)), for the unique TF ϱ that satisfies the TF equation (1.4b, 1.4c). $z \in \mathbb{R}_+^k \equiv \{(z_1, \dots, z_k) | z_i \geq 0\}$. The R_i are fixed and distinct.

Definitions. If f is a real valued function on $\mathbb{R}_+^k$ then:

(i) f is *weakly superadditive* (WSA) $\Leftrightarrow f(z_1 + z_2) \geq f(z_1) + f(z_2)$, $\forall z_1, z_2 \in \mathbb{R}_+^k$, such that $(z_1, z_2) = 0$, i.e. $(z_1)_i (z_2)_i = 0$, $\forall i$.

(ii) f is *superadditive* (SA) $\Leftrightarrow f(z_1 + z_2) \geq f(z_1) + f(z_2)$, $\forall z_1, z_2 \in \mathbb{R}_+^k$.

(iii) f is *strongly superadditive* (SSA) $\Leftrightarrow f(z_1 + z_2 + z_3) - f(z_1 + z_2) - f(z_1 + z_3) + f(z_1) \geq 0$, $\forall z_1, z_2, z_3 \in \mathbb{R}_+^k$.

(iv) f is *ray convex* $\Leftrightarrow f(\lambda z_1 + (1 - \lambda)z_2) \leq \lambda f(z_1) + (1 - \lambda)f(z_2)$, $\forall \lambda \in [0, 1]$, $z_1, z_2 \in \mathbb{R}_+^k$ and either $z_1 - z_2 \in \mathbb{R}_+^k$ or $z_2 - z_1 \in \mathbb{R}_+^k$.

(v) f is *ray concave* $\Leftrightarrow -f$ is ray convex.

(vi) f is *increasing* $\Leftrightarrow f(z_1 + z_2) \geq f(z_1)$, $\forall z_1, z_2 \in \mathbb{R}_+^k$.

Obviously,

$$f \text{ is SSA and } f(0) \leq 0 \Rightarrow f \text{ is SA} \Rightarrow f \text{ is WSA.} \quad (2.1)$$

Further relations among these definitions are proved in Appendix A. These are the following ($C^p(\mathbb{R}_+^k)$ denotes the p -fold continuously differentiable functions and subscripts denote partial derivatives):

Lemma 2.1. (i) If $f \in C^2(\mathbb{R}_+^k)$ then f is SSA $\Leftrightarrow f_{ij} \geq 0$, $\forall i, j$.

(ii) If $f \in C^1(\mathbb{R}_+^k)$ then f is SSA $\Leftrightarrow f_i$ is increasing.

Lemma 2.2. (i) If $f \in C^2(\mathbb{R}_+^k)$, $f(0)=0$, and $f_{ij} \geq 0 \forall i \neq j$, then f is WSA.

(ii) If $f \in C^1(\mathbb{R}_+^k)$, $f(0)=0$ and f_i is an increasing function of z_j for $j \neq i$, then f is WSA.

Remark. The converse implication is false as the WSA function $f(\underline{z}) = z_1 \sin^2(z_2)$ on $\mathbb{R}_+^2$ shows.

Lemma 2.3. f is SSA implies f is weakly ray-convex, i.e. f satisfies definition (iv) with $\lambda = 1/2$.

Remark. The converse implication is false, even if SSA is replaced by WSA. On $\mathbb{R}_+^2$, $f(\underline{z}) = |z_1 - z_2| + z_1 + z_2$ is convex (not merely ray-convex), increasing and $f(0)=0$, but $f(1, 1) < f(1, 0) + f(0, 1)$.

Lemma 2.4. If f is ray-convex and $f \in C^1(\mathbb{R}_+^k)$ then f_i is increasing.

Corollary 2.5. If f is ray-convex and $f \in C^1(\mathbb{R}_+^k)$ then

$$\sum_{i=1}^k z'_i f_i(\underline{z}) \leq f(\underline{z} + \underline{z}') - f(\underline{z}) \leq \sum_{i=1}^k z'_i f_i(\underline{z} + \underline{z}').$$

The theorems to be proved can now be stated.

Properties of $\phi(z, x)$ (neutral case) :

Theorem 2.6. For each fixed $x \in \mathbb{R}^3$, different from $R_1, \dots, R_k$, $\underline{z} \mapsto \phi(\underline{z}, x)$ is in $C^1(\mathbb{R}_+^k)$ and $C^2(\mathbb{R}_+^k \setminus \{0\})$. $\underline{z} \mapsto \phi_i(\underline{z}, x)$ and $\underline{z} \mapsto \phi_{ij}(\underline{z}, x)$ are equicontinuous in x . Furthermore,

(i) $\phi_{ij}(\underline{z}, x) \leq 0$, $\forall i, j$, and is negative semidefinite as a matrix, i.e. $\sum_{i,j=1}^k \bar{c}_i c_j \phi_{ij}(\underline{z}, x) \leq 0$ for all $\underline{c} \in C^k$.

(ii) $\phi_{ij}(\underline{z}, R_p) = \lim_{x \rightarrow R_p} \phi_{ij}(\underline{z}, x)$ exists ($\underline{z} \neq 0$).

(iii) $\underline{z} \mapsto \phi_i(\underline{z}, x) \geq 0$ and is ray-convex, $\forall i$.

(iv) $\lim_{x \rightarrow R_i} \{\phi_i(\underline{z}, x) - |x - R_i|^{-1}\} \leq 0$ exists. $\phi_i(\underline{z}, x) < |x - R_i|^{-1}$.

(v) $\phi_i(\underline{z}, R_j) = \lim_{x \rightarrow R_j} \phi_i(\underline{z}, x)$ exists for $i \neq j$. Moreover, $\phi_i(\underline{z}, R_j) = \phi_j(\underline{z}, R_i)$.

(vi) For every $\alpha < (1 + \sqrt{73})/2$, there exist an $R(\alpha) < \infty$ and finite numbers $M(\alpha)$ and $B(\alpha)$ such that $\phi_i(\underline{z}, x) \leq M(\alpha)|x|^{-\alpha}$, $-\phi_{ij}(\underline{z}, x) \leq B(\alpha)|x|^{-\alpha}$, ($\underline{z} \neq 0$), hold when $|x| > R(\alpha)$.

Using Lemma 2.1, we have

Corollary 2.7. For each fixed $x \in \mathbb{R}^3$, different from $R_1, R_2, \dots, R_k$,

(i) $-\phi(\underline{z}, x)$ is SSA.

(ii) $\phi(\underline{z}, x)$ is concave (not merely ray-concave).

(iii) $-\phi_i(\underline{z}, x)$ and $(\phi(\underline{z}, x))$ are strictly increasing.

Remark. That $\phi(\underline{z}, x)$ is increasing is Teller's Lemma [5], [Theorem V.5, [1]].

Properties of K , A , R and e (neutral case) :

Theorem 2.8. $K(\underline{z}) \in C^1(\mathbb{R}_+^k)$ and $C^2(\mathbb{R}_+^k \setminus \{0\})$ and :

$$(i) \quad K_i(\underline{z}) = 3 \lim_{x \rightarrow R_i} \left[\phi(\underline{z}, x) - \sum_{j=1}^k z_j \phi_j(\underline{z}, x) \right] \quad (2.2)$$

$$(ii) \quad K_{ij}(\underline{z}) = -3 \sum_{p=1}^k z_p \phi_{ij}(\underline{z}, R_p). \quad (2.3)$$

Remark. The limit in (2.2) exists by Theorem 2.6, and by $e_i = \lim_{x \rightarrow R_i} \{\phi(\underline{z}, x) - z_i|x - R_i|^{-1}\}$ ([1], Theorem II.16, Lemma V.7).

Using Theorem 2.6 we have

Corollary 2.9. (i) $K_{ij}(\underline{z}) \geq 0$ and is positive semidefinite as a matrix,

(ii) $K(\underline{z})$ is convex (not merely ray-convex) and SSA on $\mathbb{R}_+^k$,

(iii) $K(0) = 0$, which implies $K(\underline{z})$ is SA.

Theorem 2.10. (i) $R(\underline{z})$ and $A(\underline{z})$ are convex (not merely ray convex) and SSA on $\mathbb{R}_+^k$.

(ii) $e(\underline{z})$ is WSA on $\mathbb{R}_+^k$.

Remark. (ii) is just Teller's Theorem [5], [1, Theorem V.1], $e(\underline{z})$ is not SA. For $k=1$, $e(\underline{z}) = -(\text{const.})z^{7/3}$, [1], and this is not SA. However we make the following.

Conjecture 2. Let $\tilde{e}(\underline{z}) = \sum_{j=1}^k e^{at}(z_j)$, where $e^{at}(z)$ is the TF energy of an isolated atom of charge z . Then $e(\underline{z}) - \tilde{e}(\underline{z})$ is SA.

Remark. $e - \tilde{e}$ is not SSA because

$$(\partial^2 / \partial z_i^2)(e - \tilde{e}) = \lim_{x \rightarrow R_i} (\phi_i(\underline{z}, x) - (\partial \phi^{at} / \partial z)(z_i, x)),$$

and this is negative if some $z_j \neq 0$ ($j \neq i$) by Corollary 2.7 (iii). It is obvious that $e - \tilde{e}$ is WSA since e and $-\tilde{e}$ are both WSA.

$$\text{Definition. } X(\underline{z}) \equiv 3K(\underline{z}) - \sum_{i=1}^k z_i K_i(\underline{z}). \quad (2.4)$$

Theorem 2.11. $X(\underline{z})$ is SSA and $X(0) = 0$. Hence X is SA. Moreover $X(\underline{z})$ is ray convex (as follows from Lemma 2.3 and Theorem 2.8).

These theorems can be combined to yield the desired results about the pressure and compressibility.

Theorem 2.12. For the neutral molecule, the pressure and compressibility as given by (1.6), (1.7) exist and satisfy :

$$(i) \quad 3l^3 P(\underline{z}) = e(\underline{z}) + K(\underline{z}), \quad (2.5)$$

$$(ii) \quad 9l^3 \kappa^{-1}(\underline{z}) = 6l^3 P(\underline{z}) + 2e(\underline{z}) + 3X(\underline{z}) \quad (\text{cf. (2.4)}), \quad (2.6)$$

(iii) P and κ^{-1} are WSA and non-negative,

(iv) $l^2 P(\underline{z}, l)$ is a decreasing function of l . Equivalently, $e(\underline{z}, l)$ is a convex function of l .

Proof. We can write (1.8) in the form $e(\underline{z}, l) = l^{-7} e(l^3 \underline{z})$, where $e(l^3 \underline{z}) \equiv e(l^3 \underline{z}, 1)$. Hence $3l^0 P = 7e - 3l^3 \sum_{i=1}^k z_i e_i$ since $z_i e_i$ exists [1]. (1.12), (1.13) are true [1], and so (2.5) is proved [cf. (1.14)]. Since e and K are WSA, so is $e + K$ and $e + K \geq (\text{sum of the } e + K \text{ for isolated atoms}) = 0$. Using scaling again on the right side of (2.5) ($K(\underline{z}, l) = l^{-7} K(l^3 \underline{z}, 1)$ also), and Theorem 2.8(i), we can differentiate (2.5). Again using (1.13) and rescaling, (2.6) is obtained. By Theorem 2.11, (iii) is true. To prove (iv) note that for an atom, $e = -(\text{const.})z^{7/3}$ and $K = -e$; hence $2e + 3X = 0$ in this case. Since e and X are WSA, $2e + 3X \geq 0$. Thus $-l \frac{\partial P}{\partial l} \geq 2P$. If one writes $P(\underline{z}, l) = l^{-7} \pi(\underline{z}, l)$, then $\partial \pi / \partial l \leq 0$. $\square$

The following conjecture, if true, would show that $l^4 P$ is decreasing, for the right side of (2.6) is $12l^3 P(\underline{z}) + \tilde{X}(\underline{z})$. It would also show that $K(\underline{z}, l)$ is decreasing in l .

Conjecture 3. $\tilde{X}(\underline{z}) \equiv 3X(\underline{z}) - 2K(\underline{z})$ is WSA.

Remark. $\tilde{X}(\underline{z}) = 0$ for an atom.

Let us define $E(\underline{z}) = e(\underline{z}) - U(\underline{z})$. It has been proved ([1], Theorem V.3) that $-E(\underline{z})$ is WSA. We conjecture that something stronger holds, namely

Conjecture 4. $E(\underline{z}, l)$ is monotone increasing in l , for fixed $\underline{z}$.

Remark. It is easy to check that Conjecture 4 is implied by Conjecture 1. Conjecture 4 means that the pressure of a molecule in which the *nuclear-nuclear repulsion is neglected* is negative instead of positive. Some results in this direction for the Schrödinger theory are given in [11].

III. Heuristic Proofs

In this section we give simple, but non-rigorous proofs that $K(\underline{z})$ and $X(\underline{z})$, (2.4), are SSA and $K(\underline{z})$ is convex. From this, Theorem 2.12 on the positivity of P and K follows, as mentioned in Sect. II. We think it is important to provide these “proofs” because the main line of the argument may be obscure in the proper proofs given in the next section. These “proofs” assume that all necessary derivatives exist. Let us begin with some facts about the TF potential $\phi(\underline{z}, x)$. Hereafter we refer only to the neutral case. By (1.4) $\phi(\underline{z}, x)$ satisfies the TF equation

$$-(4\pi)^{-1} \Delta \phi(\underline{z}, x) + \phi(\underline{z}, x)^{3/2} = \sum_{i=1}^k z_i \delta(x - R_i). \quad (3.1)$$

The kernel for $[-(4\pi)^{-1} \Delta + \phi^{1/2}]^{-1}$ is positive, and $z_i \delta(x - R_i)$ are positive “functions”. Therefore $\phi(\underline{z}, x) \geq 0$ all x . Differentiating (3.1) twice with respect to the z 's we formally get

$$[-(4\pi)^{-1} \Delta + (3/2)\phi(\underline{z}, x)^{1/2}] \phi_i(\underline{z}, x) = \delta(x - R_i) \quad (3.2a)$$

and

$$[-(4\pi)^{-1} \Delta + (3/2)\phi(\underline{z}, x)^{1/2}] \phi_{ij}(\underline{z}, x) = -(3/4)\phi(\underline{z}, x)^{-1/2} \phi_i(\underline{z}, x) \phi_j(\underline{z}, x). \quad (3.2b)$$

From Eq. (3.2a) we have $\phi_i \geq 0$, since the kernel for $[-(4\pi)^{-1}\Delta + (3/2)\phi^{1/2}]^{-1}$ is positive. For the same reason, $\phi_{ij} \leq 0$, all i, j , and therefore $-\phi$ is SSA. Multiplying (3.2b) by $\bar{c}_i, c_j$, with $c_j \in C$, and summing over i, j we get,

$$[-(4\pi)^{-1}\Delta + (3/2)\phi(\underline{z}, x)^{1/2}] \sum_{i,j=1}^k \bar{c}_i \phi_{ij} c_j = -(3/4)\phi^{-1/2} \left| \sum_{i=1}^k c_i \phi_i \right|^2.$$

Therefore the quadratic form $\sum_{i,j=1}^k \bar{c}_i \phi_{ij}(\underline{z}, x) c_j$ is non-positive for all $\underline{c} \in C^k$. Hence ϕ is concave in $\mathbb{R}_+^k$. Finally differentiating (3.2b) with respect to z_l we have,

$$\begin{aligned} & [-(4\pi)^{-1}\Delta + (3/2)\phi^{1/2}] \phi_{ijl} \\ &= (3/8)\phi^{-3/2} \phi_i \phi_j \phi_l - (3/4)\phi^{-1/2} [\phi_{il} \phi_j + \phi_{ij} \phi_l + \phi_{jl} \phi_i], \end{aligned} \quad (3.2c)$$

which in turn implies $\phi_{ijl} \geq 0$ all i, j, l . Indeed the following is formally true: $(-1)^{n+1} \phi_{i_1 i_2 \dots i_n} \geq 0$ for all i_j and all $n \geq 1$.

Remark. If one assumes that the derivative $\phi_{i_1 \dots i_n}$ exists, then Theorem 4 of Ref. [3] shows that the sign is indeed $(-1)^{n+1}$. To use Theorem 4 for this purpose it is necessary to choose $R_i = R_j$ for some $i \neq j$, but this is allowed, as explained in [3] in the paragraph after (1.6). Theorem 4 of [3] directly gives the SSA of $-\phi$ without going through Lemma 2.1. Indeed, Theorem 4 is a generalization of SSA; for example $\phi(\underline{z}_1 + \underline{z}_2 + \underline{z}_3 + \underline{z}_4, x) - \phi(\underline{z}_1 + \underline{z}_2 + \underline{z}_3, x) - \phi(\underline{z}_1 + \underline{z}_3 + \underline{z}_4, x) - \phi(\underline{z}_1 + \underline{z}_2 + \underline{z}_4, x) + \phi(\underline{z}_1 + \underline{z}_2, x) + \phi(\underline{z}_1 + \underline{z}_3, x) + \phi(\underline{z}_1 + \underline{z}_4, x) - \phi(\underline{z}_1, x) \geq 0$. However, we are obliged to prove the existence of the first two derivatives of $\phi(\underline{z}, x)$ because we need them in our proof that $K(\underline{z})$ and $X(\underline{z})$ are SSA.

From $\phi_{ijl} \geq 0$ follows the ray-convexity of ϕ_i because the quadratic form $\sum_{j,l=1}^k (\phi_i)_{jl} z_j z_l$ is non-negative for all $\underline{z} \in \mathbb{R}_+^k$.

Now, let us formally show that K is SSA. We have to prove that $K_{ij} \geq 0$ all i, j (see Lemma 2.1). For the neutral molecule the kinetic energy is given by

$$K(\underline{z}) = (3/5) \int \phi(\underline{z}, x)^{5/2} dx. \quad (3.3)$$

Differentiating (3.3) twice with respect to the z 's we get,

$$\begin{aligned} K_{ij}(\underline{z}) &= (3/2) \left[\int \phi(\underline{z}, x)^{3/2} \phi_{ij}(\underline{z}, x) dx \right. \\ &\quad \left. + (3/2) \int \phi(\underline{z}, x)^{1/2} \phi_i(\underline{z}, x) \phi_j(\underline{z}, x) dx \right]. \end{aligned} \quad (3.4)$$

Introducing (3.2b) in (3.4), partial integration yields

$$K_{ij} = 3 \int \phi_{ij} ((4\pi)^{-1} \Delta \phi - \phi^{3/2}) dx = -3 \sum_{l=1}^k z_l \phi_{ij}(R_l),$$

where the last equality is a consequence of (3.1). But $\phi_{ij} \leq 0$, all i, j and therefore $K_{ij} \geq 0$ and K is SSA. Furthermore $[\phi_{ij}]$ is negative semi-definite [recall the discussion after (3.2b)]; hence $[K_{ij}]$ is positive semi-definite and K is convex.

It remains to be shown that $X_{ij} \geq 0$, all i, j , that is, X is SSA (Lemma 2.1). Differentiating X twice with respect to z_i, z_j we have

$$\begin{aligned} X_{ij} &= K_{ij} - \sum_{l=1}^k z_l K_{ijl} \\ &= 3 \sum_{l,m=1}^k z_l z_m \phi_{ijm}(R_l), \end{aligned}$$

where the last equality follows from (3.5). Therefore $X_{ij} \geq 0$, all i, j because $\phi_{ijm}(x) \geq 0 \forall x, i, j, m$.

IV. Proof of Theorems 2.6–2.11

Here we give the rigorous proofs of the theorems enunciated in Sect. II. Only neutral systems are considered. Let us begin by recalling some of the known facts about the TF potential $\phi(\underline{z}, x)$ that we are going to need in our proofs:

(P-1) $\phi(\underline{z}, x)$ satisfies,

$$\phi(\underline{z}, x) = \sum_{i=1}^k z_i |x - R_i|^{-1} - \int |x - y|^{-1} \phi(\underline{z}, y)^{3/2} dy. \quad (4.1)$$

(P-2) $\phi(\underline{z}, x)$ is bounded and continuous on any open subset of $\mathbb{R}^3$ which is at non-zero distance from all the R_i ([1], Theorem IV.1). In fact, the TF potential is real analytic away from all the R_j on all of $\mathbb{R}^3$ ([1], Theorem IV.6).
 $\left(\phi(\underline{z}, x) - \sum_{i=1}^k z_i |x - R_i|^{-1} \text{ is continuous for all } x. \right)$

(P-3) $\phi(\underline{z}, x)$ is strictly positive for $\underline{z} \neq 0$ ([1], Theorem IV.3).

(P-4) $|x|^4 \phi(\underline{z}, x) \rightarrow 9\pi^{-2}$ as $|x| \rightarrow \infty$, uniformly with respect to direction. (This is Sommerfeld's formula, [1], Theorem IV.10.) Moreover, for every $c < 3\pi^{-1} \exists R(c) < \infty$ such that $\phi(\underline{z}, x) \geq c^2 |x|^{-4}$ when $|x| \geq R(c)$ ([1], Theorems IV.8, IV.10).

(P-5) Properties (P-1) and (P-2) imply that $\phi(\underline{z}, x) = z_j |x - R_j|^{-1} + g(x)$ near R_j , where g is a continuous function.

(P-6) By the foregoing $\phi(\underline{z}, x) \in L^p$ for every $p \in [1, 3)$, and $\phi(\underline{z}, x)^{1/4} \in L^p$ for every $p \in (3, 12)$.

(P-7) $\phi(\underline{z}, x)$ is increasing in $\underline{z}$ for every $x \in \mathbb{R}^3$. (This is Teller's lemma, [1], Theorem V.5.)

(P-8) $\phi(\underline{z}, x)$ is strongly subadditive in $\underline{z}$ for every $x \in \mathbb{R}^3$ ([3], Theorem 4). In particular $\phi(\underline{z}, x)$ is subadditive.

(P-9) $\phi(\underline{z}, x)$ is concave in $\underline{z}$.

Proof of (P-9). Let $\psi(x) \equiv (\phi(\underline{z}, x) - \alpha \phi(\underline{z}_1, x) - (1 - \alpha) \phi(\underline{z}_2, x))$, with $\underline{z} \equiv \alpha \underline{z}_1 + (1 - \alpha) \underline{z}_2$, $0 \leq \alpha \leq 1$. By (P-2) ψ is continuous for all x and by (P-4) ψ goes to zero at infinity, hence $S = \{x | \psi(x) < 0\}$ is open and $\psi = 0$ on $\partial S \cup \{\infty\}$. On S , $-(4\pi)^{-1} \Delta \psi = -\phi(\underline{z}, x)^{3/2} + \alpha \phi(\underline{z}_1, x)^{3/2} + (1 - \alpha) \phi(\underline{z}_2, x)^{3/2} \geq -\phi(\underline{z}, x)^{3/2} + \phi(\underline{z}, x)^{3/2} = 0$ because $t \mapsto t^{3/2}$ is convex. "Hence ψ is superharmonic on S and thus ψ takes its minimum on $\partial S \cup \{\infty\}$ where it is zero. Then S is empty." $\square$

Remark. Since the argument between the apostrophes in the last paragraph repeats several times throughout this paper we will denote it by MMP (maximum modulus principle) to abbreviate.

If we call $\phi_i(z, x)$ the derivative of $\phi(z, x)$ with respect to z_i , we have formally

$$\phi_i(z, x) = |x - R_i|^{-1} - \frac{3}{2} \int |x - y|^{-1} \phi(z, y)^{1/2} \phi_i(z, y) dy. \quad (4.2)$$

Our first task will be to investigate the general properties of equations like (4.2).

IV.1. General Properties of an Integral Equation [Eq. (4.4)]

We deal here with L^p spaces ($\mathbb{R}^3$ always being understood) and with the weak L^p_w spaces:

Definition. $f \in L^p_w$ ($p > 0$) if and only if there is a constant $c < \infty$ such that $D_f(a) \equiv \mu\{x | |f(x)| > a\} \leq c^p a^{-p}$, all $a > 0$, where μ is Lebesgue measure. The infimum of all such c is denoted by $\|f\|_{p,w}$.

Remarks. (1) $L^p \subset L^p_w$ (for $p \geq 0$). If $f \in L^p$, $\|f\|_{p,w} \leq \|f\|_p$.

(2) $\|f\|_{p,w}$ is not a norm since it fails to satisfy the triangle inequality.

(3) It can be easily checked from the definition that $|x|^{-1} \in L^3_w$.

We will need the following later.

Lemma 4.1. If $f, g \in L^p_w$ and L^q_w respectively then $f \cdot g \in L^r_w$ and $\|f \cdot g\|_{r,w} \leq 2^{1/r} \|f\|_{p,w} \|g\|_{q,w}$, with $r^{-1} = p^{-1} + q^{-1}$ and $0 < p, q, r < \infty$.

Proof. Without loss we can assume $\|f\|_{p,w} = \|g\|_{q,w} = 1$.

$$\{x | |f(x)g(x)| > a\} \subset \{x | |f(x)| > a^{r/p}\} \cup \{x | |g(x)| > a^{r/q}\}.$$

Therefore $D_{fg}(a) \leq D_f(a^{r/p}) + D_g(a^{r/q})$. But $D_f(a^{r/p}) \leq a^{-r} \|f\|_{p,w}^p$ and $D_g(a^{r/q}) \leq a^{-r} \|g\|_{q,w}^q$, whence $D_{fg}(a) \leq 2a^{-r}$. $\square$

Notes. (i) The constant $2^{1/r}$ is not the best possible. It is easy to find a better one, namely $p^{1/p} q^{1/q} / r^{1/r}$.

(ii) For more details about L^p_w spaces the reader can consult [6].

The main tool to show existence and uniqueness of solutions to equations like (4.2) in some function spaces is given by:

Theorem 4.2. Let $w \in L^3_w(\mathbb{R}^3)$, w real, and let $\frac{3}{2} < p < 3$. Then the map

$$T_w : g \mapsto w(x) \int |x - y|^{-1} w(y) g(y) dy, \quad (4.3)$$

is a bounded map from $L^p(\mathbb{R}^3) \rightarrow L^p(\mathbb{R}^3)$.

Note. Theorems of this kind have been proved by Faris [9] and Strichartz [10].

Proof. By the previous lemma $A_w : g \mapsto wg$ is a bounded map from $L^p_w \rightarrow L^r_w$ with $r^{-1} = p^{-1} + 1/3$. Also A_w restricted to L^p is a bounded map by Remark (1). Now, $B : h \mapsto |x|^{-1} * h$ is a bounded map from $L^r_w \rightarrow L^t_w$ with $1 + t^{-1} = r^{-1} + (1/3)$ (since $|x|^{-1} \in L^3_w$, and the weak form of Young's inequality, [6]), when $t > 1$, $1 < r < 3/2$. Therefore $T_w = A_w B A_w$ is a bounded map from $L^p_w \rightarrow L^p_w$ for all $p \in (3/2, 3)$. Finally by the Marcinkiewicz-Zygmund interpolation theorem T_w extends to a bounded map from $L^p \rightarrow L^p$, $3/2 < p < 3$. $\square$

If we restrict the domain of T_w to L^2 , T_w is a bounded operator from the Hilbert space L^2 into itself. Moreover T_w is self-adjoint and positive since the kernel $|x-y|^{-1}$ is positive definite. Hence we have,

Corollary 4.3. *The equation $(T_w + 1)g = u$ with T_w defined by (4.3) and $w \in L_w^3$, $u \in L^2$ has a unique L^2 solution, g .*

We now obtain the main result of this section :

Theorem 4.4. *Let $w \in L_w^3$ and $wv \in L^2$. Then there is a unique f (defined a.e.) which satisfies the equation*

$$f(x) = v(x) - \int |x-y|^{-1} w(y)^2 f(y) dy \quad (4.4)$$

a.e. and such that $wf \in L^2$.

Proof. (i) Existence: Define

$$f(x) \equiv v(x) - \int |x-y|^{-1} w(y) g(y) dy, \quad (4.5)$$

where g is the unique L^2 solution (by Corollary 4.3) to

$$(T_w + 1)g = vw. \quad (4.6)$$

From (4.5) and (4.6) we have $wf = g$ and therefore the f defined in this way satisfies (4.4) and also $wf \in L^2$. (ii) Uniqueness: Assume there are two solutions f_1, f_2 to Eq. (4.4) such that wf_1 and $wf_2 \in L^2$. Both wf_1 and wf_2 satisfy Eq. (4.6) which has a unique solution. Therefore $wf_1 = wf_2$ a.e., and hence $f_1 = f_2$ a.e. [using (4.4)]. $\square$

Having shown the existence of a unique solution to (4.4) for some class of v and w , we next specialize to the particular v of interest.

Theorem 4.5. *Let $u \in \mathbb{R}^3$ and let f_u be the solution to the equation*

$$f_u(x) = |x-u|^{-1} - \int |x-y|^{-1} w(y)^2 f_u(y) dy, \quad (4.7)$$

with $w \in L_w^3$ and $h_u(x) \equiv w(x)|x-u|^{-1} \in L^2(\mathbb{R}^3)$ for all $u \in \mathbb{R}^3$. Then the integral in (4.7) is finite for all $x \in \mathbb{R}^3$ and thus $f_u(x)$ is defined by the right side of (4.7) for all $x \neq u$. Furthermore, if $u \neq t \in \mathbb{R}^3$, $f_u(t) = f_t(u)$.

Proof. Since $w \in L_w^3$ and $h_u(x) \in L^2$. Theorem 4.4 implies that there is a unique (a.e.) function solving (4.7) and satisfying $wf_u \equiv g_u \in L^2$. Since $wf_u \in L^2$ and $h_x(y) \in L^2$ for all x , the integrand in (4.7) is absolutely integrable for all x . Now, g_u satisfies $(1 + T_w)g_u = h_u$ and $(g, h_u) = (g, (1 + T_w)g_u) = ((1 + T_w)g, g_u) = (h, g_u)$ since T_w [defined in (4.3)] is self-adjoint. Explicitly, this says

$$\int w(x)^2 [f_t(x)|x-u|^{-1} - f_u(x)|x-t|^{-1}] dx = 0.$$

Using (4.7) this implies $0 = f_u(t) - |t-u|^{-1} - f_t(u) + |u-t|^{-1}$. $\square$

Up to now the only assumption on w was $w \in L_w^3$. We will now make a stronger assumption about w in order to obtain continuity of the solution to (4.7). First, a preliminary remark :

Lemma 4.6. *If $f \in L^p$ and $g \in L^q$ with p, q dual indices different from 1 and ∞ then $f * g$ is a bounded continuous function going to zero at infinity.*

Proof. This result is standard. See [1], Lemma II.25. $\square$

Lemma 4.7. *Let $w \in L_w^3$ and such that $w \in L^{6-\varepsilon} \cap L^{6+\varepsilon}$ (for some $\varepsilon > 0$). Let v be such that $vw \in L^2$ and let f denote the solution to (4.4). Then the integral in (4.4) (namely $f - v$) is a bounded continuous function going to zero at infinity.*

Proof. By Theorem 4.4 the solution f exists and satisfies $g \equiv wf \in L^2$. Using Hölder's inequality, $wg \in L^{p^-} \cap L^{p^+}$ with $p_{\pm} = 2(6 \pm \varepsilon)(8 \pm \varepsilon)^{-1}$. We can always decompose $|x|^{-1} = |x|_{<}^{-1} + |x|_{>}^{-1}$ with $|x|_{<}^{-1} \in L^{3-\eta_+}$, $|x|_{>}^{-1} \in L^{3+\eta_-}$ ($\eta_{\pm}$, positive). Choose $\eta_{\pm} = \varepsilon(4 \pm \varepsilon)^{-1}$; then $(3 \mp \eta_{\pm})$ is the dual of $p_{\pm}$. But $f - v = |x|^{-1} * gw = |x|_{<}^{-1} * gw + |x|_{>}^{-1} * gw$, and hence this lemma follows from Lemma 4.6. $\square$

Remark. The w which will eventually be used is simply $\phi^{1/4}(\underline{z}, x)$. This satisfies the conditions of Lemma 4.7 by (P-6).

We now study the dependence of the solution f on v and w .

Lemma 4.8. *Let $w \in L_w^3$ and such that $w \in L^{6-\varepsilon} \cap L^{6+\varepsilon}$ (for some $\varepsilon > 0$). Let $u \in \mathbb{R}^3$ be a parameter, and let*

$$v_u(x) = |x - u|^{-1} + V(x), \quad (4.8)$$

where $V(x)$ is a continuous superharmonic function, bounded and going to zero at infinity such that $wV \in L^2$, then:

- (i) *The solution f_u to (4.4) is non-negative for all x .*
- (ii) *If v_u is fixed and if $w_1(x)^2 \geq w_2(x)^2$ all x , the corresponding solutions f_1 , (resp. f_2) to (4.4) with $w = w_1$ (resp. $w = w_2$) are such that $f_1(x) \leq f_2(x)$ all x .*
- (iii) *Now keep w fixed. Let v_{1u}, v_{2u} be of the form (4.8) with $v_{1u} - v_{2u}$ superharmonic, then the corresponding solutions f_{1u}, f_{2u} are such that $f_{1u}(x) \geq f_{2u}(x)$ all x .*

Proof. Since $w \in L^{6-\varepsilon} \cap L^{6+\varepsilon}$ and $|x - u|^{-1} \in L^{3-\eta_+} + L^{3+\eta_-}$ with $\eta_{\pm} = \varepsilon(4 \pm \varepsilon)^{-1}$, using Hölder's inequality we have $w(x)|x - u|^{-1} \in L^2$. Therefore $v_u w \in L^2$ and, by Theorem 4.4, there is a unique solution f_u to Eq. (4.4), with this v_u satisfying $wf_u \in L^2$. Moreover by Lemma 4.6 and the properties of v_u, f_u [defined by the right side of (4.4)] is continuous away from u and goes to zero at infinity. (i) Let $S = \{x | f(x) < 0\}$. Since $f_u \rightarrow \infty$ as $x \rightarrow u$, S is disjoint from u and open (since f_u is continuous away from u). On S , the distributional laplacian of f_u is given by

$$-(4\pi)^{-1} \Delta f_u = -w^2 f_u - (4\pi)^{-1} \Delta V \geq -w^2 f_u \geq 0.$$

Then (i) follows from MMP. (ii) Call $\psi = f_2 - f_1$. ψ is continuous everywhere and goes to zero at infinity. Let $S = \{x | \psi(x) < 0\}$. S is open and $\psi = 0$ on $\partial S \cup \{\infty\}$. On S , $-(4\pi)^{-1} \Delta \psi = w_1^2 f_1 - w_2^2 f_2 \geq -w_2^2 \psi > 0$ and (ii) follows using MMP. (iii) is a consequence of (i) and the linearity of f_u in v_u . $\square$

Theorem 4.9. *(Asymptotic Behavior of $f(x)$). Consider $v(x) = |x|^{-1}$ and w as in Lemma 4.8 and, moreover, $w(x)^2 \geq c|x|^{-2}$ for $|x| > R$ and some $c > 0$. Then $f(x) \leq M(c)|x|^{-\alpha(c)}$ for $|x| > R$ where $\alpha(c) = (1 + \sqrt{1 + 16\pi c})/2$ and $M(c) = \alpha(c)^{-1} R^{\alpha(c)-1}$.*

Proof. Take w_1 defined by $w_1(x)^2 = c|x|^{-2}$ for $|x| > R$ and $w_1^2 = 0$ for $|x| < R$, the solution to (4.4) corresponding to this w is given by $f_1(x) = M(c)|x|^{-\alpha(c)}$ for $|x| > R$ and $f_1(x) = |x|^{-1}(1 + a|x|)$ for $|x| < R$ with $M(c) = \alpha(c)^{-1}R^{\alpha(c)-1}$ and $a = (1 - \alpha(c))/(\alpha(c)R)$. We have $w(x)^2 \geq w_1(x)^2$ all x and hence, by Lemma 4.8(ii) $f(x) \leq f_1(x)$. In particular $f(x) \leq M(c)|x|^{-\alpha(c)}$ for $|x| > R$. $\square$

To close this section we prove the following,

Lemma 4.10. *If $v(x) = |x|^{-1}$, and w as in Lemma 4.8, then $\int w(x)^2 f(x) dx \leq 1$, If, moreover, $w(x)^2 \geq c|x|^{-2}$ for $|x| > R$ (for some c, R), then $\int w(x)^2 f(x) dx = 1$.*

Proof. Assume $\int w^2 f > 1$. Define the spherical average $[f](r) \equiv (4\pi)^{-1} \int_{S_2} f(r\Omega) d\Omega$. From (4.4) we have $[f](r) = r^{-1} - \int w(y)^2 f(y) [(4\pi)^{-1} \int_{S_2} d\Omega |r\Omega - y|^{-1}] dy$. Using the well known formula $(4\pi)^{-1} \int_{S_2} d\Omega |r\Omega - y|^{-1} = \{\max(r, |y|)\}^{-1}$ we get $[f](r) \leq r^{-1}(1 - \int_{|y| \leq r} w(y)^2 f(y) dy)$. Therefore for r large enough $[f](r) < 0$ which contradicts Lemma 4.8(i). Hence $\int w^2 f \leq 1$. Let us now consider w such that $w(x)^2 \geq c|x|^{-2}$ for $|x| > R$. $[f](r) \geq r^{-1}(1 - \int w^2 f)$ by the same arguments as above. If $\int w^2 f < 1$, then $[f](r) \geq dr^{-1}$ for some positive d which contradicts Theorem 4.9. $\square$

IV.2. Proof of Theorem 2.6: Properties of the TF Potential

The strategy to prove that $\underline{z} \rightarrow \phi(\underline{z}, x) \in C^1$ is the following: we first show that a unique solution to Eq. (4.2) exists (Lemma 4.11) and is continuous in $\underline{z}$ uniformly with respect to x (Lemma 4.13). We then show that $\phi_i^{\varepsilon}(\underline{z}, x) \equiv \varepsilon^{-1}[\phi(\underline{z} + \varepsilon \underline{e}_i, x) - \phi(\underline{z}, x)]$, with $\underline{e}_i = (\delta_i^j)$ a unit vector in $\mathbb{R}_+^k$ along z_i , converges to $\phi_{,i}(\underline{z}, x)$ as $\varepsilon \rightarrow 0$ uniformly in x . (Lemma 4.14). We then imitate the same argument to show that $\phi \in C^2$.

In what follows we study the equation

$$\phi_u(\underline{z}, x) = |x - u|^{-1} - (3/2) \int dy |x - y|^{-1} \phi(\underline{z}, y)^{1/2} \phi_u(\underline{z}, y). \quad (4.9)$$

Note that $w \equiv (3/2)^{1/2} \phi^{1/4} \in L_w^3$ (since w goes as $|x|^{-1}$ at infinity) and $w \in L^p$ for all $p \in (3, 12)$ [(P-4), (P-6)]. In particular $w \in L^{6-\varepsilon} \cap L^{6+\varepsilon}$, for some $\varepsilon > 0$, therefore $|x - u|^{-1} w \in L^2$ as discussed in the proof of Lemma 4.8.

Lemma 4.11. *(Existence of $\phi_u(\underline{z}, x)$). There is a unique $\phi_u(\underline{z}, x)$ satisfying Eq. (4.9) with $\phi_u \phi^{1/4} \in L^2$, and it has the following properties:*

- (i) $\phi_u(\underline{z}, x) - |x - u|^{-1}$ is a bounded continuous function going to zero at infinity.
- (ii) $\underline{z} \rightarrow \phi_u(\underline{z}, x)$ is non-negative and decreasing
- (iii) $\underline{z} \rightarrow \phi_u(\underline{z}, x)$ is ray-convex.
- (iv) For every $\alpha < (1 + \sqrt{73})/2 \simeq 4.77$, there exists an $R(\alpha) < \infty$ and a finite number $M(\alpha)$ such that $\phi_u(\underline{z}, x) \leq M(\alpha)|x|^{-\alpha}$ for $|x| \geq R(\alpha)$.

Proof. Since $w \in L_w^3$ and $|x - u|^{-1} w \in L^2$, Theorem 4.4 implies the existence of a unique $\phi_u(\underline{z}, x)$ satisfying (4.9) with $w\phi_u \in L^2$. (i) follows from Lemma 4.7, since

$w \in L^{6-\epsilon} \cap L^{6+\epsilon}$. As for (ii), Lemma 4.8(i) implies that $\underline{z} \mapsto \phi_u(\underline{z}, x) \geq 0$ all x ; (P-7) together with Lemma 4.8(ii) imply that $\underline{z} \mapsto \phi_u(\underline{z}, x)$ is decreasing. To prove (iii), let $\underline{z}_1, \underline{z}_2 \in \mathbb{R}_+^k$ with $\underline{z}_1 - \underline{z}_2 \in \mathbb{R}_+^k$ and define $\underline{z} = \lambda \underline{z}_1 + (1 - \lambda) \underline{z}_2$ with $0 \leq \lambda \leq 1$. Define $\psi(x) \equiv \lambda \phi_u(\underline{z}_1, x) + (1 - \lambda) \phi_u(\underline{z}_2, x) - \phi_u(\underline{z}, x)$. Because of (i) $\psi(x)$ is continuous everywhere and goes to zero at infinity. Then $S = \{x | \psi(x) < 0\}$ is open and $\psi = 0$ on $\partial S \cup \{\infty\}$. From (4.9)

$$-(4\pi)^{-1} \Delta \psi = \frac{3}{2} \{ -\lambda \phi(\underline{z}_1, x)^{1/2} \phi_u(\underline{z}_1, x) - (1 - \lambda) \phi(\underline{z}_2, x)^{1/2} \phi_u(\underline{z}_2, x) \\ + \phi(\underline{z}, x)^{1/2} \phi_u(\underline{z}, x) \}.$$

Because of (P-9), (P-7) and part (ii) (since $\underline{z}_1 - \underline{z}_2 \in \mathbb{R}_+^k$) we have $-(4\pi)^{-1} \Delta \psi \geq 0$. Hence (iii) follows using MMP. (iv) given $\alpha < (1 + \sqrt{73})/2$ (i.e. $c < (9/2\pi)$) there exists $R(c) < \infty$ such that $w(x)^2 \equiv (3/2) \phi(\underline{z}, x)^{1/2} \geq c|x|^{-\alpha}$ (P-4). Hence, by Theorem 4.9, $\phi_u(\underline{z}, x) \leq M(\alpha)|x|^{-\alpha}$ for $|x| \geq R(c)$. $\square$

Remark. In the atomic case, Hille [7] used methods of ordinary differential equations to prove that the asymptotic formula with $\alpha = (1 + \sqrt{73})/2$ was exact [[7], Eq. (4.5)].

We now prove a general theorem that we will later need:

Theorem 4.12. *Let f be a real (or complex) function on $\mathbb{R}_+^k$. Suppose f satisfies the following condition:*

$$|f(\underline{z}_1) - f(\underline{z}_2)| < K \|\underline{z}_1 - \underline{z}_2\|_2^\alpha, \quad (4.10)$$

for all $\underline{z}_1, \underline{z}_2 \in \mathbb{R}_+^k$ such that $\underline{z}_1 - \underline{z}_2 \in \mathbb{R}_+^k$, for some $\alpha > 0$ and some $K > 0$, then $\underline{z} \mapsto f(\underline{z})$ is continuous in the whole of $\mathbb{R}_+^k$.

Proof. Assume first that $\underline{z} \in \text{Int}(\mathbb{R}_+^k)$. Let $\underline{n} = (1, 1, \dots, 1)$, and $\underline{z}_0 \equiv \underline{z} - \delta \underline{n}$, with $\delta \leq \min_{1 \leq i \leq k} (z_i)$ (i.e. $\underline{z}_0 \in \mathbb{R}_+^k$). Let $\underline{z}' \in B(\underline{z}, \delta)$, the ball of radius δ centered at $\underline{z}$. Applying (4.10) twice we get $|f(\underline{z}') - f(\underline{z})| \leq (\|\underline{z}' - \underline{z}_0\|_2^\alpha + \|\underline{z} - \underline{z}_0\|_2^\alpha) K$, because $\underline{z}' - \underline{z}_0 \in \mathbb{R}_+^k$, $\underline{z} - \underline{z}_0 \in \mathbb{R}_+^k$. But, as $\delta \rightarrow 0$, $\|\underline{z} - \underline{z}_0\|_2 \rightarrow 0$ and $\|\underline{z}' - \underline{z}_0\| \rightarrow 0$ uniformly in $B(\underline{z}, \delta)$, so f is continuous at $\underline{z}$. Now, if $\underline{z}$ is in one face, F , of $\mathbb{R}_+^k$ (of dimension $0 \leq l < k$) the same argument can be repeated using $\tilde{\underline{n}} \equiv$ projection of $\underline{n}$ on F . $\square$

Lemma 4.13. *(Continuity of $\phi_u(\underline{z}, x)$ in $\underline{z}$). $\phi_u(\underline{z}, x)$ defined as the solution to (4.9) (Satisfying $\phi^{1/4} \phi_u \in L^2$) is continuous for all $\underline{z} \in \mathbb{R}_+^k$ uniformly with respect to x .*

Proof. We divide the proof into two steps. First we prove continuity at $\underline{z} \neq \underline{0}$, and then at $\underline{z} = \underline{0}$. (i) $\underline{z} \neq \underline{0}$. There is a $\underline{z}^* \in \mathbb{R}_+^k$, $\underline{z}^* \neq \underline{0}$ such that $\underline{z} - \underline{z}^* \in \mathbb{R}_+^k$. Let $\underline{z}_1, \underline{z}_2 \in (\underline{z}^* + \mathbb{R}_+^k)$ with $\underline{z}_1 - \underline{z}_2 \in \mathbb{R}_+^k$. From Eq. (4.9) we get,

$$\phi_u(\underline{z}_2, x) - \phi_u(\underline{z}_1, x) = (3/2) \int |x - y|^{-1} \phi_u(\underline{z}_1, y) [\phi(\underline{z}_1, y)^{1/2} - \phi(\underline{z}_2, y)^{1/2}] dy \\ + (3/2) \int |x - y|^{-1} \phi(\underline{z}_2, y)^{1/2} [\phi_u(\underline{z}_1, y) - \phi_u(\underline{z}_2, y)]. \quad (4.11)$$

Since $\underline{z}_1 - \underline{z}_2 \in \mathbb{R}_+^k$, $\phi_u(\underline{z}_2, x) - \phi_u(\underline{z}_1, x) \geq 0$ because of Lemma 4.11(ii). Hence, (4.11) implies

$$|\phi_u(\underline{z}_2, x) - \phi_u(\underline{z}_1, x)| \leq (3/2) \int |x - y|^{-1} \phi_u(\underline{z}^*, y) (\phi(\underline{z}_1, y)^{1/2} - \phi(\underline{z}_2, y)^{1/2}).$$

Because of (P-8) we have,

$$\phi(\underline{z}_1, y)^{1/2} - \phi(\underline{z}_2, y)^{1/2} \leq \phi(\underline{z}_1 - \underline{z}_2, y)^{1/2} \leq \|\underline{z}_1 - \underline{z}_2\|_2^{1/2} \sum_{i=1}^k \frac{1}{|y - R_i|^{1/2}}, \quad (4.12)$$

where the last inequality follows from Eq. (4.1). Hence

$$|\phi_u(\underline{z}_2, x) - \phi_u(\underline{z}_1, x)| < \|\underline{z}_1 - \underline{z}_2\|_2^{1/2} g(x), \quad (4.13a)$$

where

$$g(x) = (3/2) \int |x - y|^{-1} \sum_{i=1}^k \left(\frac{1}{|y - R_i|^{1/2}} \right) \phi_u(\underline{z}^*, y). \quad (4.13b)$$

By Young's inequality $g(x) \in L^\infty$ because $|x|^{-1} \in L^4 + L^{5/2}$ and $|y - R_i|^{-1/2} \phi_u(\underline{z}^*, y) \in L^p$ for any $1 \leq p < 2$, in particular for $p = 4/3$ and $p = 5/3$. [Lemma 4.11(i), (iv)]. Theorem 4.12 and Eq. (4.13) then imply that $\phi_u(\underline{z}, x)$ is continuous in $\underline{z}$, uniformly with respect to x , for all $\underline{z} \in (\underline{z}^* + \mathbb{R}_+^k)$. But $\bigcup_{\underline{z}^*} (\underline{z}^* + \mathbb{R}_+^k) = \mathbb{R}_+^k \setminus \{\underline{0}\}$. (ii) $\underline{z} = \underline{0}$, Equation (4.9) and Lemma 4.11(ii) imply

$$|\phi_u(\underline{z}, x) - \phi_u(\underline{0}, x)| \leq h(x) \equiv \int |x - y|^{-1} \phi(\underline{z}, y)^{1/2} |y - u|^{-1} dy. \quad (4.14)$$

Using Young's inequality we get,

$$\begin{aligned} \|h(x)\|_\infty &\leq \|\phi(\underline{z}, \cdot)^{1/4}\|_6 (\| |x|^{-1} \|_{42/17} \|\phi^{1/4} |y - u|^{-1}\|_{7/3} \\ &\quad + \| |x|^{-1} \|_4 \|\phi^{1/4} |y - u|^{-1}\|_{12/7}) \end{aligned}$$

and thus $\|h(x)\|_\infty \leq c \|\phi^{1/4}\|_6$, with $c < \infty$ because $\phi^{1/4} |y - u|^{-1} \in L^p$ for any $(3/2) < p < (12/5)$ and $|x|^{-1} \in L^p$ ($p < 3$), $|x|^{-1} \in L^p$ ($p > 3$). From (4.14) we finally get $\|\phi_u(\underline{z}, x) - \phi_u(\underline{0}, x)\|_\infty \leq c \|\phi^{1/4}\|_6 = c(\int \varrho)^{1/6} = cz^{1/6}$, where $z = \sum_{i=1}^k z_i$. $\square$

Let us define $\phi_i(\underline{z}, x)$ to be $\phi_u(\underline{z}, x)$ with $u = R_i$. Then the last step to prove that $\phi(\underline{z}, x) \in C^1(\mathbb{R}_+^k)$, uniformly with respect to x , is the following:

Lemma 4.14. (Convergence of $\phi_i^e(\underline{z}, x)$ to $\phi_i(\underline{z}, x)$). Let $\phi_i^e(\underline{z}, x) \equiv \varepsilon^{-1} [\phi(\underline{z} + \varepsilon \underline{e}_i, x) - \phi(\underline{z}, x)]$ with $\underline{e}_i = (\delta_i^1)$ being a unit vector in $\mathbb{R}_+^k$, along z_i and $\varepsilon \geq -z_i$. Then $\phi_i^e(\underline{z}, x) \rightarrow \phi_i(\underline{z}, x)$ as $\varepsilon \rightarrow 0$, uniformly with respect to x .

Proof. (i) Consider first $\varepsilon > 0$. We will prove the following,

$$\phi_i^e(\underline{z}, x) - \phi_i(\underline{z} + \varepsilon \underline{e}_i, x) \geq 0, \quad (4.15a)$$

$$\phi_i^e(\underline{z}, x) - \phi_i(\underline{z}, x) \leq 0. \quad (4.15b)$$

Consider $\psi(x) \equiv \phi_i^e(\underline{z}, x) - \phi_i(\underline{z} + \varepsilon \underline{e}_i, x)$. By Lemma 4.11 and (P-2), ψ is continuous for all x and goes to zero at infinity. Then $S \equiv \{x | \psi < 0\}$ is open and $\psi = 0$ on $\partial S \cup \{\infty\}$. On S ,

$$\begin{aligned} -(4\pi)^{-1} \Delta \psi &= \varepsilon^{-1} [\phi(\underline{z}, x)^{3/2} - \phi(\underline{z} + \varepsilon \underline{e}_i, x)^{3/2}] \\ &\quad + (3/2) \phi_i(\underline{z} + \varepsilon \underline{e}_i, x) \cdot \phi(\underline{z} + \varepsilon \underline{e}_i, x)^{1/2} \\ &\geq (2\varepsilon)^{-1} \phi(\underline{z} + \varepsilon \underline{e}_i, x)^{1/2} \phi(\underline{z}, x) [\mu^2 - 3 + 2\mu^{-1}], \end{aligned}$$

where $\mu^2 \equiv \phi(\underline{z} + \varepsilon \underline{e}_i, x) / \phi(\underline{z}, x) \geq 1$ [by (P-7)]. Hence $\mu^2 - 3 + 2\mu^{-1} \geq 0$. MMP then implies (4.15a). The proof of (4.15b) is analogous. From (4.15) and Lemma 4.13, $\|\phi_i^{\varepsilon}(\underline{z}, x) - \phi_i(\underline{z}, x)\|_{\infty} \rightarrow 0$ as $\varepsilon \downarrow 0$. (ii) If $-z_i < \varepsilon < 0$, (4.15a, b) imply

$$\phi_i(\underline{z}, x) \leq \phi_i^{\varepsilon}(\underline{z}, x) \leq \phi_i(\underline{z} + \varepsilon \underline{e}_i, x) \quad (4.16)$$

which in turn implies $\|\phi_i^{\varepsilon}(\underline{z}, x) - \phi_i(\underline{z}, x)\|_{\infty} \rightarrow 0$ as $\varepsilon \uparrow 0$. $\square$

If we denote by $\phi_{ij}(\underline{z}, x)$ the derivative of $\phi_i(\underline{z}, x)$ with respect to z_j , we formally get (from (4.2)):

$$\begin{aligned} -\phi_{ij}(\underline{z}, x) &= (3/4) \int |x-y|^{-1} \phi(\underline{z}, y)^{-1/2} \phi_i(\underline{z}, y) \phi_j(\underline{z}, y) dy \\ &\quad - (3/2) \int |x-y|^{-1} \phi(\underline{z}, y)^{1/2} (-\phi_{ij}(\underline{z}, y)) dy. \end{aligned} \quad (4.17)$$

As we have already mentioned, the strategy to prove that $\phi(\underline{z}, x)$ is in $C^2(\mathbb{R}_+^k \setminus \underline{0})$, uniformly with respect to x , will be the same as before. Now there will be an additional difficulty, namely the control of $\phi(\underline{z}, y)^{-1/2}$. Let us start proving that a solution to (4.17) indeed exists.

Lemma 4.15. (Existence of $\phi_{ij}(\underline{z}, x)$). For $\underline{z} \neq \underline{0}$, there is a unique $\phi_{ij}(\underline{z}, x)$ satisfying Eq. (4.17) and such that $\phi_{ij}\phi^{1/4} \in L^2$. Moreover:

- (i) $\phi_{ij}(\underline{z}, x)$ is continuous for all x . It is bounded and goes to zero at infinity.
- (ii) $-\phi_{ij}(\underline{z}, x)$ is non-negative and so is $\sum_{1 \leq i, j \leq k} \bar{c}_i (-\phi_{ij}(\underline{z}, x)) c_j$, and $\underline{c} \in C^k$.
- (iii) $-\phi_{ij}(\underline{z}, x)$ is a decreasing function of $\underline{z}$.

Proof. Note first that, for $\underline{z} \neq \underline{0}$, $\phi(\underline{z}, \cdot)^{-1/4} \phi_i(\underline{z}, \cdot) \in L^q$ for any $1 \leq q < 4$. In fact, for $\underline{z} \neq \underline{0}$, ϕ is strictly positive (P-3) and $\phi(\underline{z}, x) \geq \tilde{c}|x|^{-4}$ for $|x| > R = 2 \max |R_i|$ and some positive constant $\tilde{c}$ (P-4). Because of Lemma 4.11(iv), $\phi_i \phi^{-1/4} \leq c_i |x|^{1-\alpha}$ for $|x| > R(\alpha)$ with $4 < \alpha < 4.77$. Then if $B(0, R(\alpha)) = \{x | |x| \leq R(\alpha)\}$, $\phi_i \phi^{-1/4} \in L^p(\mathbb{R}^3 \setminus B(0, R(\alpha)))$, $\forall p \geq 1$. Inside $B(0, R(\alpha))$ and away from R_i $\phi_i \phi^{-1/4} \phi_i$ is bounded since B is compact, (P-4) and Lemma 4.11(i). In a neighborhood of R_i $\phi_i \phi^{-1/4}$ behaves like $|x - R_i|^{-3/4}$, hence $\phi_i \phi^{-1/4} \in L^q$ for any $1 \leq q < 4$. Therefore, $\phi_i \phi^{-1/4} \phi_j \phi^{-1/4} \in L^s$ for $1 \leq s < 2$ and, since $|x|^{-1} \in L^3_w$, $v_{ij} \equiv (3/4)|x|^{-1} * (\phi_i \phi^{-1/4} \phi_j \phi^{-1/4}) \in L^t$ for any $3 < t < \infty$ (by the generalized Young's inequality). Moreover, $v_{ij} \phi^{1/4} \in L^2$ because $\phi^{1/4} \in L^p$, $3 < p < 12$. Finally, since $\phi^{1/4} \in L^3_w$, Theorem 4.4 implies that there is a unique solution $\phi_{ij}(\underline{z}, x)$ ($\underline{z} \neq \underline{0}$) to Eq. (4.17) satisfying $\phi^{1/4} \phi_{ij} \in L^2$. (i) By Lemma 4.7, $\phi_{ij} + v_{ij}$ is a bounded continuous function going to zero at infinity. Lemma 4.6 shows that v_{ij} is continuous, bounded and goes to zero at infinity because $(\phi^{-1/4} \phi_i)(\phi^{-1/4} \phi_j) \in L^s$ for any $1 \leq s < 2$ and $|x|^{-1} \in L^4 + L^{5/2}$. (ii) Follows from Lemma 4.8 because v_{ij} is superharmonic and so is $\sum \bar{c}_i v_{ij} c_j$. Lemma 4.11(ii) and (P-7) imply that $v_{ij}(\underline{z}_1, x) - v_{ij}(\underline{z}_2, x)$ is superharmonic if $\underline{z}_1 - \underline{z}_2 \in \mathbb{R}_+^k$. Then (iii) follows from Lemma 4.8(ii) and (iii). $\square$

In order to prove the asymptotic behavior of $\phi_{ij}(\underline{z}, x)$ we will need the following comparison Lemma. See also [[1], Theorem 4.7].

Lemma 4.16. Assume that f_1, f_2 are continuous positive functions on $\{x | |x| \geq R\}$ with the following properties:

- (i) $f_1(x), f_2(x) \rightarrow 0$ as $|x| \rightarrow \infty$.

(ii) $-(4\pi)^{-1}\Delta f_1 = \varrho_1 - w_1 f_1$, $-(4\pi)^{-1}\Delta f_2 = \varrho_2 - w_2 f_2$, where the derivatives and equalities are in distributional sense, and $\varrho_1(x) \geq \varrho_2(x)$, $0 \leq w_1(x) \leq w_2(x)$.

(iii) $f_1(x) \geq f_2(x)$ for all x such that $|x| = R$. Then $f_1(x) \geq f_2(x)$ for all x such that $|x| \geq R$.

Proof. Define $\psi = f_1 - f_2$. Let $S = \{x | \psi(x) < 0\}$, which is open. On S $-(4\pi)^{-1}\Delta\psi = (\varrho_1 - \varrho_2) - w_1\psi + f_2(w_2 - w_1) \geq 0$. The Lemma follows from MMP since, by (i) and (iii), $\psi = 0$ at ∞ and $\psi > 0$ on ∂S . $\square$

Lemma 4.17. (*Asymptotic Behavior of $\phi_{ij}(\underline{z}, x)$*). Let $\alpha < (1 + \sqrt{73})/2 \simeq 4.77$ and let $\phi_{ij}(\underline{z}, x)$ be the solution to (4.17) satisfying $\phi^{1/4}\phi_{ij} \in L^2$. Then there exists an $R(\alpha) < \infty$ such that $-\phi_{ij}(\underline{z}, x) \leq B(\alpha)|x|^{-\alpha}$ for some $B(\alpha) > 0$, when $|x| > R(\alpha)$.

Remark. The remark below Lemma 4.11 also applies to $-\phi_{ij}$.

Proof. Consider the equation,

$$-(4\pi)^{-1}\Delta f = br^{2(1-\alpha)} - dr^{-2}f, \quad \text{for } r = |x| > R \quad (4.18)$$

where $\alpha(\alpha-1) = 4\pi d$, together with the boundary condition $f(x) = N$ for $|x| = R$. The solution $f(x)$ to (4.18), going to zero at infinity is,

$$f(x) = N(R/r)^\alpha + 4\pi b(3\alpha-5)^{-1}(\alpha-4)^{-1}r^{-\alpha}R^{4-\alpha}(1-(R/r)^{\alpha-4}). \quad (4.19)$$

Given any $\alpha(c) < (1 + \sqrt{73})/2$ there exists $R(c) < \infty$ such that $w(x)^2 \equiv (3/2)\phi(\underline{z}, x)^{1/2} \geq c|x|^{-4}$ (P-4) and $\phi_i(\underline{z}, x) \leq M(\alpha(c))|x|^{-\alpha(c)}$ (Lemma 4.11(iv)) for $|x| \geq R(c)$. Hence $(3/4)\phi^{-1/2}\phi_i \leq (3/4)M(c)^2c^{-1}|x|^{2-2\alpha(c)}$ for $|x| \geq R(c)$. Using (4.17), (4.18) and the comparison Lemma we get:

$$-\phi_{ij}(\underline{z}, x) \leq [3\pi M^2c^{-1}(3\alpha-5)^{-1}(\alpha-4)^{-1}]r^{-\alpha}R^{4-\alpha}(1-(R/r)^{\alpha-4}) + N(R/r)^\alpha$$

for $|x| \geq R$, with $N = \max_{|x|=R} (-\phi_{ij}(\underline{z}, x))$ which is finite because ϕ_{ij} is bounded (Lemma 4.15(i)). $\square$

Lemma 4.18. (*Continuity of $\phi_{ij}(\underline{z}, x)$*). $\phi_{ij}(\underline{z}, x)$ is continuous in $\underline{z}$ for all $\underline{z} \in \mathbb{R}_+^k \setminus \{0\}$, uniformly in x .

Proof. Let $\underline{z}^* \in \mathbb{R}_+^k \setminus \{0\}$ such that $\underline{z} - \underline{z}^* \in \mathbb{R}_+^k$. Let $\underline{z}_1, \underline{z}_2 \in (\underline{z}^* + \mathbb{R}_+^k)$ with $(\underline{z}_1 - \underline{z}_2) \in \mathbb{R}_+^k$. Lemma 4.15(iii) and Eq. (4.17) imply

$$0 \leq \phi_{ij}(\underline{z}_1, x) - \phi_{ij}(\underline{z}_2, x) \leq I(x) + J(x), \quad (4.20a)$$

with

$$\begin{aligned} I(x) = & (3/4) \int |x-y|^{-1} \{ \phi(\underline{z}_2, y)^{-1/2} \phi_i(\underline{z}_2, y) \phi_j(\underline{z}_2, y) \\ & - \phi(\underline{z}_1, y)^{-1/2} \phi_i(\underline{z}_1, y) \phi_j(\underline{z}_1, y) \} dy \end{aligned} \quad (4.20b)$$

and

$$J(x) = (3/2) \int |x-y|^{-1} (-\phi_{ij}(\underline{z}^*, y)) [\phi(\underline{z}_1, y)^{1/2} - \phi(\underline{z}_2, y)^{1/2}] dy. \quad (4.20c)$$

To estimate $J(x)$ we use (4.12) to get.

$$J(x) \leq \|\underline{z}_1 - \underline{z}_2\|_2^{1/2} g_1(x), \quad (4.21)$$

where $g_1(x) = (3/2) \int |x-y|^{-1} (-\phi_{ij}(z^*, y)) \sum_{n=1}^k |y-R_n|^{-1/2} dy$. Since for $z \neq 0$ ϕ_{ij} is bounded everywhere [Lemma 4.15(i)] and since it goes to zero at least as fast as $r^{-4.5}$ at infinity (Lemma 4.17), $|y|^{-1/2} \phi_{ij} \in L^p$ for any $1 \leq p < 6$. In particular $|y-R_n|^{-1/2} \phi_{ij} \in L^{4/3} \cap L^2$ and therefore, by Young's inequality $g_1 \in L^\infty$ because $|x|^{-1} \in L^4 + L^2$. $I(x)$ is decomposed as,

$$I(x) = I_1(x) + I_2(x) + I_3(x) \quad (4.22)$$

where

$$I_1(x) = (3/4) \int |x-y|^{-1} (\phi(z_2, y)^{-1/2} - \phi(z_1, y)^{-1/2}) \phi_i(z_2, y) \phi_j(z_2, y) dy, \quad (4.23a)$$

$$I_2(x) = (3/4) \int |x-y|^{-1} (\phi_i(z_2, y) - \phi_i(z_1, y)) \phi(z_2, y)^{-1/2} \phi_j(z_2, y) dy, \quad (4.23b)$$

$$I_3(x) = (3/4) \int |x-y|^{-1} (\phi_j(z_2, y) - \phi_j(z_1, y)) \phi(z_1, y)^{-1/2} \phi_i(z_1, y) dy. \quad (4.23c)$$

To find a bound for $I_1(x)$ we use Lemma 4.11(ii) and the following estimate:

$$\begin{aligned} & \phi(z_2, y)^{-1/2} - \phi(z_1, y)^{-1/2} \\ &= [\phi(z_1, y) - \phi(z_2, y)] \phi(z_1, y)^{-1} \phi(z_2, y)^{-1} [\phi(z_1, y)^{-1/2} + \phi(z_2, y)^{-1/2}]^{-1} \\ &\leq \|z_1 - z_2\|_2 \left[\sum_{i=1}^k |x - R_i|^{-1} \right] (1/2) \phi(z^*, y)^{-3/2}. \end{aligned}$$

which follows from Eq. (4.1) and (P-7). Hence we have

$$I_1(x) \leq \|z_1 - z_2\|_2 g_2(x), \quad (4.24)$$

with $g_2(x) = (3/8) \int |x-y|^{-1} \phi(z^*, y)^{-3/2} \phi_i(z^*, y) \phi_j(z^*, y) \sum_{n=1}^k |x-R_n|^{-1} dy$. Since $\phi_i \phi^{-1} \in L^p$ for any $p \geq 6$, $(\phi_i \phi^{-1})(\phi_j \phi^{-1}) \in L^p$, $p \geq 3$. Also, $\phi^{1/2} \sum_{n=1}^k |x-R_n|^{-1} \in L^q$, $1 < q < 2$ and therefore $\phi^{-3/2} \phi_i \phi_j \sum_{n=1}^k |x-R_n|^{-1} \in L^s$, $1 \leq s < 2$. Since $|x|^{-1} \in L^4 + L^{5/2}$ $g_2 \in L^\infty$ by Young's inequality. Equations (4.12) and (4.23) and Lemma (4.13) imply

$$I_2(x) \leq (3/4) \int |x-y|^{-1} g(y) \|z_1 - z_2\|_2^{1/2} \phi(z^*, y)^{-1/2} \phi_j(z^*, y) dy$$

with

$$g(x) = (3/2) \int |x-y|^{-1} \sum_{i=1}^k |y-R_i|^{-1/2} \phi_i(z^*, y) dy \in L^\infty$$

because of Lemma 4.13. Then

$$I_2(x) \leq \|z_1 - z_2\|_2^{1/2} g_3(x) \quad (4.25)$$

with

$$g_3(x) = (3/4) \|g\|_\infty \int |x-y|^{-1} \phi(z^*, y)^{-1/2} \phi_j(z^*, y) dy.$$

Since $\phi^{-1/2}\phi_j \in L^p$ for any $(3/2.5) \leq p < 6$ [Lemma 4.11(iv) and (P-4)] and $|x|^{-1} \in L^4 + L^2$ we have $g_3 \in L^\infty$. Then the lemma follows from Eqs. (4.20a), (4.21), (4.24), (4.25) and Theorem 4.12. $\square$

We conclude with the proof that $\phi \in C^2(\mathbb{R}_+^k \setminus \mathcal{Q})$ uniformly in x , with the following

Lemma 4.19. (*Convergence of $\phi_{ij}^\varepsilon(\underline{z}, x)$ to $\phi_{ij}(\underline{z}, x)$*): Let

$$\phi_{ij}^\varepsilon(\underline{z}, x) \equiv \varepsilon^{-1} [\phi_i(\underline{z} + \varepsilon \underline{e}_j, x) - \phi_i(\underline{z}, x)]$$

with $\underline{e}_j = [\delta_j^i]$ unit vector in $\mathbb{R}_+^k$ along z_j and $\varepsilon \geq -z_j$. Then $\phi_{ij}^\varepsilon(\underline{z}, x) \rightarrow \phi_{ij}(\underline{z}, x)$ as $\varepsilon \rightarrow 0$, uniformly in x .

Proof. (i) Consider first $\varepsilon > 0$. As in Lemma 4.14 we prove first:

$$\phi_{ij}(\underline{z}, x) \leq \phi_{ij}^\varepsilon(\underline{z}, x) \leq \phi_{ij}(\underline{z} + \varepsilon \underline{e}_j, x). \quad (4.26)$$

Let $\psi(x) = \phi_{ij}^\varepsilon(\underline{z}, x) - \phi_{ij}(\underline{z}, x)$. By Lemmas 4.11(i) and 4.15(i), ψ is continuous everywhere and goes to zero at infinity. Then $S = \{x | \psi < 0\}$ is open and $\psi = 0$ on $\partial S \cup \{\infty\}$. Since $\varepsilon > 0$, using Lemma 4.11(ii) we get,

$$\begin{aligned} -(4\pi)^{-1} \Delta \psi &\geq -(3/2) \phi(\underline{z}, x)^{1/2} \psi \\ &+ (3/2\varepsilon) \phi_i(\underline{z} + \varepsilon \underline{e}_j, x) [\phi(\underline{z}, x)^{1/2} - \phi(\underline{z} + \varepsilon \underline{e}_j, x)^{1/2} + (\varepsilon/2) \phi(\underline{z}, x)^{-1/2} \phi_j(\underline{z}, x)]. \end{aligned}$$

Moreover, $\phi(\underline{z}, x)^{1/2} - \phi(\underline{z} + \varepsilon \underline{e}_j, x)^{1/2} + (1/2) \phi(\underline{z}, x)^{-1/2} \phi_j(\underline{z}, x) \geq 0$ because $\phi^{1/2}(\underline{z}, x)$ is concave (P-9), and ϕ is $C^2(\mathbb{R}_+^k)$ for each x . Therefore, on S $-(4\pi)^{-1} \Delta \psi \geq 0$ and by the MMP the first inequality in (4.26) follows. The other one is proved in the same way. Lemma 4.18 and (4.26) then imply $\|\phi_{ij}^\varepsilon(\underline{z}, x) - \phi_{ij}(\underline{z}, x)\|_\infty \rightarrow 0$ as $\varepsilon \downarrow 0$ for $\underline{z} \neq \mathcal{Q}$.

(ii) If $-z_j < \varepsilon < 0$, (4.26) is replaced by $\phi_{ij}(\underline{z} + \varepsilon \underline{e}_j, x) \leq \phi_{ij}^\varepsilon(\underline{z}, x) \leq \phi_{ij}(\underline{z}, x)$ and the lemma follows from that. $\square$

IV.3. Proof of Theorems 2.8–2.11: Properties of K, A, R, e , and X

We begin by proving that $K(\underline{z})$ is in $C^1(\mathbb{R}_+^k)$ and $C^2(\mathbb{R}_+^k \setminus \mathcal{Q})$.

Lemma 4.20. (*Existence of $K_i(\underline{z})$*). Let $K(\underline{z}) \equiv (3/5) \int \phi(\underline{z}, x)^{5/2} dx$. Then

$$K_i(\underline{z}) \equiv \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} [K(\underline{z} + \varepsilon \underline{e}_i) - K(\underline{z})]$$

exists and is equal to $(3/2) \int \phi(\underline{z}, x)^{3/2} \phi_i(\underline{z}, x) dx$, where $\underline{e}_i = [\delta_i^j]$ is a unit vector along z_i .

Proof. (i) Consider first the case $\varepsilon \geq 0$. Then

$$\varepsilon^{-1} [K(\underline{z} + \varepsilon \underline{e}_i) - K(\underline{z})] = (3/5) \int \varepsilon^{-1} [\phi(\underline{z} + \varepsilon \underline{e}_i, x)^{5/2} - \phi(\underline{z}, x)^{5/2}] dx. \quad (4.27)$$

Now,

$$\varepsilon^{-1} [\phi(\underline{z} + \varepsilon \underline{e}_i, x)^{5/2} - \phi(\underline{z}, x)^{5/2}] = \varepsilon^{-1} P(\mu) [\phi(\underline{z} + \varepsilon \underline{e}_i, x) - \phi(\underline{z}, x)] \phi(\underline{z} + \varepsilon \underline{e}_i, x)^{3/2}, \quad (4.28)$$

where $P(\mu) = (1 + \mu^{5/2})^{-1}(1 + \mu + \mu^2 + \mu^3 + \mu^4)$ and $\mu \equiv \phi(\underline{z}, x)\phi(\underline{z} + \varepsilon \underline{e}_i, x)^{-1} < 1$ by (P-7). Hence $P(\mu) \leq 5/2$. Moreover, because of (P-9) and Theorem 2.6,

$$\phi(\underline{z} + \varepsilon \underline{e}_i, x) - \phi(\underline{z}, x) \leq \varepsilon \phi_i(\underline{z}, x) \quad (4.29)$$

Using (4.29), (4.27), and (4.28) we get:

$$\begin{aligned} \varepsilon^{-1} [\phi(\underline{z} + \varepsilon \underline{e}_i, x)^{5/2} - \phi(\underline{z}, x)^{5/2}] &\leq (5/2) \phi_i(\underline{z}, x) \phi(\underline{z} + \varepsilon \underline{e}_i, x)^{3/2} \\ &\leq (5/2) |x - R_i|^{-1} \phi(\underline{z} + \underline{e}_i, x)^{3/2}, \end{aligned}$$

where the last inequality follows from (P-7) (assuming $\varepsilon \leq 1$) and Lemma 4.11(ii). Since $\phi^{3/2}(\underline{z}, x) \in L^1 \cap L^2$ at least, $|x - R_i|^{-1} \phi^{3/2}(\underline{z} + \underline{e}_i, x) \in L^1$. Hence the lemma follows by Theorem 2.6 and dominated convergence. (ii) In the case $-z_i \leq \varepsilon \leq 0$, an analysis similar to the above yields

$$\begin{aligned} \varepsilon^{-1} [\phi(\underline{z} + \varepsilon \underline{e}_i, x)^{5/2} - \phi(\underline{z}, x)^{5/2}] &\leq 5/2 \phi_i(\underline{z} + \varepsilon \underline{e}_i, x) \phi(\underline{z}, x)^{3/2} \\ &\leq (5/2) |x - R_i|^{-1} \phi(\underline{z}, x)^{3/2} \in L^1. \quad \square \end{aligned}$$

Lemma 4.20 assures us that the derivatives of $K(\underline{z})$ along the axis exist. The proof that $K(\underline{z})$ is in fact in $C^1(\mathbb{R}_+^k)$ is provided by the following:

Lemma 4.21. (*Continuity of $K_i(\underline{z})$*). $K_i(\underline{z}) \equiv (3/2) \int \phi(\underline{z}, x)^{3/2} \phi_i(\underline{z}, x) dx$ is continuous for all $\underline{z} \in \mathbb{R}_+^k$.

Proof. Let us prove continuity at $\underline{z}$. Let $\underline{z}^* = 2 \max_{1 \leq i \leq k} (z_i)(1, 1, \dots, 1)$. Let $A \equiv \{\underline{z}' \in \mathbb{R}_+^k \mid \underline{z}^* - \underline{z}' \in \mathbb{R}_+^k\}$. Obviously $\underline{z} \in A$. Consider $\underline{z}' \in A$, $\underline{z}' \rightarrow \underline{z}$. By Corollary 2.7(iii),

$$\phi(\underline{z}', x)^{3/2} \phi_i(\underline{z}', x) \leq |x - R_i|^{-1} \phi(\underline{z}^*, x)^{3/2} \in L^1.$$

Moreover, as $\underline{z}' \rightarrow \underline{z}$, $\phi(\underline{z}', x)^{3/2} \phi_i(\underline{z}', x) \rightarrow \phi(\underline{z}, x)^{3/2} \phi_i(\underline{z}, x)$ everywhere (Theorem 2.6). The lemma follows by dominated convergence. $\square$

The following is a useful alternative expression for $K_i(\underline{z})$.

Lemma 4.22.

$$K_i(\underline{z}) = 3 \lim_{x \rightarrow R_i} \left\{ \phi(\underline{z}, x) - \sum_{j=1}^k z_j \phi_j(\underline{z}, x) \right\}. \quad (4.30)$$

Note. Because of (P-3) and Theorem 2.6(iv), the above limit exists.

Proof. From Eqs. (4.1) and (4.2) we have

$$\begin{aligned} F(\underline{z}, x) &\equiv \phi(\underline{z}, x) - \sum_{j=1}^k z_j \phi_j(\underline{z}, x) \\ &= (3/2) \sum_{j=1}^k \int |x - y|^{-1} \phi(\underline{z}, y)^{1/2} z_j \phi_j(\underline{z}, y) dy - \int |x - y|^{-1} \phi(\underline{z}, y)^{3/2} dy. \end{aligned} \quad (4.31)$$

The two integrals on the right side of (4.31) are bounded and continuous everywhere, therefore,

$$F(\underline{z}, R_i) = (3/2) \sum_{j=1}^k \int |y - R_i|^{-1} z_j \phi_j(\underline{z}, y) \phi(\underline{z}, y)^{1/2} dy - \int |y - R_i|^{-1} \phi(\underline{z}, y)^{3/2} dy. \quad (4.32)$$

Because of Eq. (4.2) and Theorem 2.6(v) we have

$$\int |y - R_i|^{-1} \phi(\underline{z}, y)^{1/2} \phi_j(\underline{z}, y) dy = \int |y - R_j|^{-1} \phi(\underline{z}, y)^{1/2} \phi_i(\underline{z}, y) dy \quad (4.33)$$

for all i, j . From Eqs. (4.32) and (4.33) we get,

$$F(\underline{z}, R_i) = (3/2) \int \phi(\underline{z}, y)^{1/2} \phi_i(\underline{z}, y) \sum_{j=1}^k z_j |y - R_j|^{-1} dy - \int |y - R_i|^{-1} \phi(\underline{z}, y)^{3/2} dy.$$

Combining the last equation and Eqs. (4.1) and (4.2) we finally get

$$F(\underline{z}, R_i) = (1/2) \int \phi(\underline{z}, y)^{3/2} \phi_i(\underline{z}, y) dy = K_i(\underline{z})/3,$$

which is Eq. (4.30). Note that to get the first equality we have used

$$\begin{aligned} & \int dy \phi(\underline{z}, y)^{1/2} \phi_i(\underline{z}, y) \left[\int dw |w - y|^{-1} \phi(\underline{z}, w)^{3/2} \right] \\ &= \int dy \phi(\underline{z}, y)^{3/2} \left[\int dw |w - y|^{-1} \phi(\underline{z}, w)^{1/2} \phi_i(\underline{z}, w) \right], \end{aligned}$$

which is true by Fubini's theorem, since $\phi(\underline{z}, y)^{1/2} \phi_i(\underline{z}, y) \in L^1$ (Lemma 4.10) and $\int dw |w - y|^{-1} \phi(\underline{z}, w)^{3/2} \in L^\infty$ (Theorem IV.1, [1]). $\square$

The right side of (4.30) can be written in terms of the right sides of the integral Eqs. (4.1) and (4.2). Using the same kind of dominated convergence argument as in the proof of Lemmas 4.20 and 4.21, it is easy to check that K_i is differentiable and

$$\begin{aligned} K_{ij}(\underline{z}) &= -3 \sum_{l=1}^k z_l \phi_{lj}(\underline{z}, R_l) \\ &= -3 \sum_{l=1}^k z_l \phi_{ij}(\underline{z}, R_l), \end{aligned} \quad (4.34)$$

where the last equality follows by using Theorem 2.6.

Proof of Theorem 2.8. $K \in C^1$, $K \in C^2(\mathbb{R}^k \setminus \mathbb{Q})$ follow from Lemmas 4.20, 4.21, Eq. (4.34) and Theorem 2.6. (i) is proved in Lemma 4.22 and (ii) is Eq. (4.34). $\square$

Proof of Theorem 2.10. (i) Let us start with the convexity of $R(\underline{z})$. Define $\underline{z} \equiv \alpha \underline{z}_1 + (1 - \alpha) \underline{z}_2$, $\alpha \in [0, 1]$. Consider now the following identity:

$$\begin{aligned} & 2[R(\underline{z}) - \alpha R(\underline{z}_1) - (1 - \alpha) R(\underline{z}_2)] \\ &= 2 \int dx dy [\phi(\underline{z}, x)^{3/2} - \alpha \phi(\underline{z}_1, x)^{3/2} - (1 - \alpha) \phi(\underline{z}_2, x)^{3/2}] |x - y|^{-1} \phi(\underline{z}, y)^{3/2} \\ &\quad - \alpha \int dx dy [\phi(\underline{z}, x)^{3/2} - \phi(\underline{z}_1, x)^{3/2}] |x - y|^{-1} [\phi(\underline{z}, y)^{3/2} - \phi(\underline{z}_1, y)^{3/2}] \\ &\quad - (1 - \alpha) \int dx dy [\phi(\underline{z}, x)^{3/2} - \phi(\underline{z}_2, x)^{3/2}] |x - y|^{-1} [\phi(\underline{z}, y)^{3/2} - \phi(\underline{z}_2, y)^{3/2}]. \end{aligned} \quad (4.35)$$

The last two terms of the left side of (4.35) are negative because $|x|^{-1}$ is a positive kernel. Moreover from Eq. (4.1) we get the following identity

$$\int dx dy |x-y|^{-1} [\phi(\underline{z}, x)^{3/2} - \alpha \phi(\underline{z}_1, x)^{3/2} - (1-\alpha) \phi(\underline{z}_2, x)^{3/2}] \phi(\underline{z}, y)^{3/2} \\ = - \int dx dy [\phi(\underline{z}, x) - \alpha \phi(\underline{z}_1, x) - (1-\alpha) \phi(\underline{z}_2, x)] \phi(\underline{z}, y)^{3/2} \leq 0, \quad (4.36)$$

where the last inequality follows from the concavity of $\phi(\underline{z}, x)$ (P-9). From (4.35) and (4.36) the convexity follows. The SSA is proved in a similar way. The Virial theorem (Theorem II.23, [1]) yields

$$A(\underline{z}) = (5/3)K(\underline{z}) + 2R(\underline{z}), \quad (4.37)$$

and hence the convexity and SSA of A follow from those of K [Corollary 2.9(ii)] and R . (ii) That $e(\underline{z})$ is WSA on $\mathbb{R}_+^k$ is proven in [1], Theorem V.7. See also [3], Theorem 1. (This is in fact Teller's Theorem [5]). $\square$

Proof of Theorem 2.11. Equation (2.4) and Theorem 2.8 imply

$$X_i(\underline{z}) = 2K_i(\underline{z}) - \sum_{j=1}^k z_j K_{ji}(\underline{z}). \quad (4.38)$$

Using now (2.2), (2.3), and (4.38) we get

$$X_i(\underline{z}) = \lim_{x \rightarrow R_i} 6N(\underline{z}, x), \quad (4.39)$$

with

$$N(\underline{z}, x) = \phi(\underline{z}, x) - \sum_{j=1}^k z_j \phi_j(\underline{z}, x) + (1/2) \sum_{j,l=1}^k z_l z_j \phi_{jl}(\underline{z}, x). \quad (4.40)$$

Note that in order to obtain (4.39) we have used $\phi \in C^2$ and also Theorem 2.6(v). Note also that the limit in (4.39) in fact exists because of (P-2), Theorem 2.6(iv) and (ii). Let us compute $N(\underline{z} + \underline{\varepsilon}, x) - N(\underline{z}, x)$ for $\underline{\varepsilon} \in \mathbb{R}_+^k$. Using (P-9) and $\phi \in C^1$ we have,

$$\phi(\underline{z} + \underline{\varepsilon}, x) - \phi(\underline{z}, x) \geq \sum_{j=1}^k \varepsilon_j \phi_j(\underline{z} + \underline{\varepsilon}, x). \quad (4.41)$$

From (4.40) and (4.41),

$$N(\underline{z} + \underline{\varepsilon}, x) - N(\underline{z}, x) \geq - \sum_{j=1}^k z_j [\phi_j(\underline{z} + \underline{\varepsilon}, x) - \phi_j(\underline{z}, x)] \\ + (1/2) \sum_{j,l=1}^k (\underline{z} + \underline{\varepsilon})_l (\underline{z} + \underline{\varepsilon})_j \phi_{jl}(\underline{z} + \underline{\varepsilon}, x) - (1/2) \sum_{l,j=1}^k z_l z_j \phi_{jl}(\underline{z}, x). \quad (4.42)$$

The ray-convexity of ϕ_i [Theorem 2.6(iii)] implies:

$$\sum_{j=1}^k \varepsilon_j \phi_{ij}(\underline{z}, x) \leq \phi_i(\underline{z} + \underline{\varepsilon}, x) - \phi_i(\underline{z}, x) \leq \sum_{j=1}^k \varepsilon_j \phi_{ij}(\underline{z} + \underline{\varepsilon}, x).$$

We conclude that

$$N(\underline{z} + \underline{\varepsilon}, x) - N(\underline{z}, x) \geq (1/2) \sum_{j,l=1}^k \varepsilon_l \varepsilon_j \phi_{jl}(\underline{z} + \underline{\varepsilon}, x), \quad (4.43)$$

since $\phi_{jl}(z + \varepsilon, x) \geq \phi_{jl}(z, x)$ [Lemma 4.15(iii)]. Although the right side of (4.43) is negative, it is second order in ε . The following Lemma 4.23 shows that under these conditions $N(z, x)$ must in fact be increasing. Hence $X_i(z)$ is increasing and by Lemma 2.1 X is SSA. $\square$

Lemma 4.23. *Suppose $f: [a, b] \rightarrow \mathbb{R}$ is a real-valued function such that for every $x \in [a, b]$ there is a $c(x)$ such that $f(z) - f(x) \geq c(x)(z - x)^2$ for all $z \in [a, b]$ with $z \geq x$. Suppose further that $c \in L^1([a, b])$. Then f is increasing, i.e. $z > x \Rightarrow f(z) \geq f(x)$.*

Proof. Let $N > 1$ be an integer and let I_j for $j = 1, \dots, n$ be the interval $I_j = (x + (j-1)(z-x)/n, x + j(z-x)/n)$. Then

$$\delta \equiv f(z) - f(x) = \sum_{j=0}^n (f(y_{j+1}) - f(y_j))$$

with $y_0 = x$, $y_{n+1} = z$ and $y_j \in I_j$. Without loss of generality we can assume $c(x) \leq 0$, all x . Then

$$\delta \geq \left\{ n^{-1} \sum_{j=0}^n c(y_j) \right\} \{4(z-x)^2/n\}$$

because $y_{j+1} - y_j \leq 2(z-x)/n$. Let $d_j = \int_{I_j} c(x) dx$ and $d = \sum_{j=1}^n d_j = \int_x^z c \geq \int_a^b c > -\infty$.

For each j , there exists $y_j \in I_j$ such that $c(y_j) \int_{I_j} 1 \geq d_j$, otherwise $\int_{I_j} c < d_j$. Using these y_j we have

$$\delta \geq \left\{ n^{-1} c(x) + n^{-1} \sum_{j=1}^n d_j ((z-x)/n)^{-1} \right\} \{4(z-x)^2/n\}.$$

Taking $n \rightarrow \infty$ proves the lemma. $\square$

Appendix A

Properties of Superadditive and Convex Functions on $\mathbb{R}_+^k$

The definition of superadditive and convex functions on $\mathbb{R}_+^k$, as well as many of their properties, were stated in Section II. Those properties, Lemmas 2.1 to 2.4 and Corollary 2.5, will be proved here.

Proof of Lemma 2.1. (i) $\Rightarrow$ is trivial because $f \in C^2(\mathbb{R}_+^k)$. To prove $\Leftarrow$, define $F(\lambda, \mu) \equiv f(\underline{z} + \lambda \underline{z}_1 + \mu \underline{z}_2) - f(\underline{z} + \lambda \underline{z}_1) - f(\underline{z} + \mu \underline{z}_2) + f(\underline{z})$. Then, for $\underline{z}_1, \underline{z}_2 \in \mathbb{R}_+^k$,

$$F_\lambda(\lambda, \mu) \equiv (\partial F(\lambda, \mu) / \partial \lambda) = \sum_{i=1}^k (z_1)_i [f_i(\underline{z} + \lambda \underline{z}_1 + \mu \underline{z}_2) - f_i(\underline{z} + \lambda \underline{z}_1)], \quad (\text{A.1})$$

and

$$F_{\lambda\mu}(\lambda, \mu) = \sum_{i,j=1}^k (z_1)_i (z_2)_j f_{ij}(\underline{z} + \lambda \underline{z}_1 + \mu \underline{z}_2) \geq 0, \quad (\text{A.2})$$

where the last inequality follows from $f_{ij} \geq 0$ all i, j . But $F_\lambda(\lambda, 0) \geq 0$, and hence $F_\lambda(\lambda, \mu) \geq 0$. Also, $F(0, \mu) = 0$, and hence $F(\lambda, \mu) \geq 0$. (ii) $\Rightarrow$ follows immediately from the definition of an SSA function and the fact that $f \in C^1(\mathbb{R}_+^k)$. To prove $\Leftarrow$ note that if $f \in C^1(\mathbb{R}_+^k)$ and f_i is increasing we get, from (A.1), $F_\lambda(\lambda, \mu) \geq 0$. But $F(0, \mu) = 0$ and therefore $F(\lambda, \mu) \geq 0$. $\square$

Proof of Lemma 2.2. This is similar to the previous one, taking into account that $f_{ij} \geq 0$ all i, j , $i \neq j$, and $z_1 \cdot z_2 = 0$ imply

$$\sum_{i,j=1}^k (z_1)_i (z_2)_j f_{ij} \geq 0. \quad \square$$

Proof of Lemma 2.3. If f is SSA, taking $z_3 = z_2$, in definition (iii) we have that $f((1/2)z_1 + (1/2)(z_1 + 2z_2)) \leq (f(z_1)/2) + (f(z_1 + 2z_2)/2)$. $\square$

Lemma 2.4 is a well known fact for differentiable convex functions. See [8], for example.

Proof of Corollary 2.5.

$$\begin{aligned} f(z_1 + z_2) - f(z_1) &= \int_0^1 \frac{d}{d\lambda} (f(z_1 + \lambda z_2)) d\lambda \\ &= \int_0^1 \sum_{i=1}^k (z_2)_i f_i(\lambda z_2 + z_1) d\lambda \leq \sum_{i=1}^k (z_2)_i f_i(z_2 + z_1) \end{aligned}$$

because f_i is increasing, by Lemma 2.4. The other inequality is proved in the same way. $\square$

Appendix B

Positivity of the Pressure Under Separation Relative to a Plane (in the Symmetric Case)

Consider $2k$ nuclei with coordinates $R_1, \dots, R_k$ and $R_{-1}, \dots, R_{-k}$ and strictly positive charges $z_1, \dots, z_k, z_{-1}, \dots, z_{-k}$ satisfying (for $i = 1, \dots, k$)

$$\begin{aligned} z_i &= z_{-i} \\ R_i^1 &= R_{-i}^1, R_i^2 = R_{-i}^2, -R_{-i}^3 = R_i^3 > 0. \end{aligned}$$

Let $e(l)$ denote the TF energy for this molecule when R_i^3 is replaced by $R_i^3 + l$, $i > 0$ and by $R_i^3 - l$, $i < 0$. The electron charge λ is immaterial but is fixed at some value

$$\lambda \leq 2 \sum_{i=1}^k z_i.$$

Theorem B.1. *The pressure is strictly positive, i.e.,*

$$e(l) < e(0) \quad \text{for } l > 0. \quad (\text{B.1})$$

Proof. The proof consists of showing that if the charge distribution (electron and nuclear) is cut in two parts at the $x^3 = 0$ plane, and then pulled apart by a distance

2l, the energy is lowered. Let $\varrho(x)$ be the TF density when $l=0$. Define $\varrho_l(x)$ by

$$\begin{aligned}\varrho_l(x) &= \varrho(x - (0, 0, l)), \quad x^3 \geq l \\ \varrho_l(x) &= \varrho(x + (0, 0, l)), \quad x^3 \leq -l \\ \varrho_l(x) &= 0, \quad -l < x^3 < l.\end{aligned}$$

Clearly $\int \varrho_l = \int \varrho = \lambda$, and we will use ϱ_l as a trial density for the l problem. We will show that $\mathcal{E}_l(\varrho_l) < \mathcal{E}(\varrho) = e(0)$, where $\mathcal{E}_l$ (resp. $\mathcal{E}$) is the energy functional for l (resp. 0). Obviously, $K(\varrho_l) = K(\varrho)$.

Let $D \subset \mathbb{R}^3$ be the domain $\{(x^1, x^2, x^3) | x^3 \geq 0\}$. For any function $f: D \rightarrow \mathbb{C}$, let $W_l(f) = \int_D \int_D dx dy \bar{f}(x) f(y) K_l(x, y)$,

$$K_l(x, y) = \{(x^1 - y^1)^2 + (x^2 - y^2)^2 + (x^3 + y^3 + 2l)^2\}^{-1/2}. \quad (\text{B.2})$$

In other words, $W_l(f)$ is the Coulomb interaction energy between a charge distribution f , supported on the $x^3 \geq 0$ side of the xy plane, and its (complex conjugate) reflection through the plane $x^3 = -l$. It is easy to see that

$$\mathcal{E}_l(\varrho_l) - \mathcal{E}(\varrho) = W_l(\mu) - W_0(\mu),$$

where μ is the charge density for $x^3 \geq 0$ for the $l=0$ problem, namely for $x^3 \geq 0$

$$\mu(x) = -\varrho(x) + \sum_{j=1}^k z_j \delta(x - R_j). \quad (\text{B.3})$$

Since $\mu \neq 0$, and $W_0(\mu) = \lim_{l \downarrow 0} W_l(\mu)$, the following Lemma B.2 completes the proof. $\square$

Lemma B.2. (*Reflection Positivity of the Coulomb Potential*). *Let f be a non-null function with support in $D = \{x | x^3 \geq 0\}$ and with $f \in L^1(D)$. Then, for $l > 0$, $W_l(f) > 0$, and $W_l(f)$ is a finite, strictly decreasing function of l . Moreover, $W_l(f)$ is a log convex function of l , vanishing at $l = \infty$.*

Proof. Using the well-known representation for $|x|^{-1}$, we have that

$$\begin{aligned}K_l(x, y) &= (2\pi^2)^{-1} \int d^3 p |p|^{-2} \exp\{i[p^1(x^1 - y^1) + p^2(x^2 - y^2) \\ &\quad + p^3(x^3 + y^3 + 2l)]\} \\ &= (2\pi)^{-1} \int d^2 p |p|^{-1} \bar{g}_p(x) g_p(y) \exp(-2|p|l)\end{aligned}$$

with $g_p(x) = \exp[ip^1 x^1 + ip^2 x^2 - |p|x^3]$.

We have used the fact that

$$\int_{-\infty}^{\infty} dp^3 [(p^3)^2 + a^2]^{-1} \exp[ip^3(x^3 + y^3 + 2l)] = (\pi/a) \exp[-a(x^3 + y^3 + 2l)]$$

when $x^3 + y^3 + 2l > 0$, as it is here. For $p \in \mathbb{R}^2$, let $h(p) = \int_D f(x) g_p(x) d^3 x$. Since $f \in L^1(D)$, $|h(p)| \leq \|f\|_1$, and $h(p)$ is null if and only if f is null. For $l > 0$ Fubini's theorem yields

$$W_l(f) = (2\pi)^{-1} \int d^2 p |p|^{-1} |h(p)|^2 \exp(-2|p|l). \quad (\text{B.4})$$

The representation (B.4) proves the lemma. $\square$

References

1. Lieb, E. H., Simon, B.: Advan. Math. **23**, 22–116 (1977)
2. Lieb, E. H., Simon, B.: Phys. Rev. Letters **31**, 681–683 (1973)
Lieb, E. H.: Proc. Int. Congress of Math., Vancouver (1974)
Lieb, E. H.: Rev. Mod. Phys. **48**, 553–569 (1976)
3. Benguria, R., Lieb, E. H.: Ann. Phys. (N.Y.) **110**, 34–45 (1978)
4. Balász, N.: Phys. Rev. **156**, 42–47 (1967)
5. Teller, E.: Rev. Mod. Phys. **34**, 627–631 (1962)
6. Reed, M., Simon, B.: Methods of modern mathematical physics. Vol. II, Fourier analysis and self-adjointness. New York: Academic Press 1975
7. Hille, E.: Proc. Nat. Acad. Sci. **62**, 7–10 (1969)
8. Rockafellar, R. T.: Convex analysis. Princeton: University Press 1970
9. Faris, W.: Duke Math. J. **43**, 365–372 (1976)
10. Strichartz, R. S.: J. Math. Mech. **16**, 1031–1060 (1967)
11. Lieb, E. H., Simon, B.: Monotonicity of the electronic contribution to the Born-Oppenheimer energy. J. Phys. B (London) **11**, L537–542 (1979)
12. Brezis, H., Lieb, E. H.: Long range atomic potentials in Thomas-Fermi theory. Commun. math. Phys. (submitted)

Communicated by J. Glimm

Received July 24, 1978

Note Added in Proof

In a recent related work [12], H. Brezis and E. H. Lieb have proved that the interaction among neutral atoms in Thomas-Fermi theory behaves, for large separation l , like l^{-7} .

Erratum

The Positivity of the Pressure in Thomas-Fermi Theory

R. Benguria and E. H. Lieb

Commun. Math. Phys. **63**, 193–218 (1978)

The conclusion of Lemma 2.4 should read :

then $\sum_{i=1}^k w_i f_i(\underline{z} + \lambda \underline{w})$ is increasing in λ for $\lambda \geq 0$ and $\underline{z}, \underline{w} \in \mathbb{R}_+^k$.

In the proof of Corollary 2.5 “... f_i is increasing” should be replaced by “... $\Sigma(z_2)_i f_i(\lambda \underline{z}_2 + \underline{z}_1)$ is increasing ...”. Lemma 2.4 was used only to prove Corollary 2.5, and the latter was used in the equation following (4.42).

The Thomas–Fermi–von Weizsäcker Theory of Atoms and Molecules^{*}

Rafael Benguria¹, Haim Brezis², and Elliott H. Lieb³

1. The Rockefeller University, New York, NY 10021, USA, on leave from Universidad de Chile, Santiago, Chile
2. Département de Mathématiques, Université Paris VI, F-75230, Paris Cedex 05, France
3. Departments of Mathematics and Physics, Princeton University, Princeton, NJ 08544, USA

Abstract. We place the Thomas–Fermi–von Weizsäcker model of atoms on a firm mathematical footing. We prove existence and uniqueness of solutions of the Thomas–Fermi–von Weizsäcker equation as well as the fact that they minimize the Thomas–Fermi–von Weizsäcker energy functional. Moreover, we prove the existence of binding for two very dissimilar atoms in the frame of this model.

Introduction

The Thomas–Fermi theory of atoms [1] (TF), attractive because of its simplicity, is not satisfactory because it yields an electron density with incorrect behavior very close and very far from the nucleus. Moreover, it does not allow for the existence of molecules. In order to correct this, von Weizsäcker [2] suggested the addition of an inhomogeneity correction

$$U_w(\rho) = C_w(\nabla\rho)^2/\rho \quad (1)$$

to the kinetic energy density. Here $c_w = h^2/(32\pi^2m)$, where m is the mass of the electron. This correction has also been obtained as the first order correction to the TF kinetic energy in a semi-classical approximation to the Hartree–Fock theory [3].

The Thomas–Fermi–von Weizsäcker (henceforth TFW) energy functional for nuclei of charges $z_i > 0$ (which need not be integral) located at $R_i, i = 1, \dots, k$ is defined by

$$\begin{aligned} \xi(\rho) = & (3\pi^2)^{-2/3} \int (\nabla\rho^{1/2}(x))^2 dx + \frac{3}{5} \int \rho(x)^{5/3} dx \\ & - \int V(x)\rho(x)dx + \frac{1}{2} \iint \rho(x)\rho(y)|x-y|^{-1} dx dy, \end{aligned} \quad (2)$$

in units in which $h^2(8m)^{-1}(3/\pi)^{2/3} = 1$ and $|e| = 1$. Here $\rho(x) \geq 0$ is the electron

^{*} Research supported by U. S. National Science Foundation under Grants MCS78-20455 (R. B.), PHY-7825390 A 01 (H. B. and E. L.), and Army Research Grant DAH 29-78-6-0127 (H. B.)

density, and

$$V(x) = \sum_{i=1}^k z_i |x - R_i|^{-1}. \quad (3)$$

While the pure TF problem has been placed on a rigorous mathematical footing [1], no parallel study has been made for the TFW problem. Such a study was undertaken in the Ph.D. thesis of one of us [4]; in this paper some of the results of [4] will be presented together with some newer results.

In this article we will study a rather more general functional, which contains the TFW energy functional (2) as a particular case. In fact, for $\rho(x) \geq 0$ and $V(x)$ given by (3), let us introduce the functional

$$\xi_p(\rho) = \int |\nabla \rho^{1/2}(x)|^2 dx + \frac{1}{p} \int \rho^p(x) dx - \int V(x) \rho(x) dx + D(\rho, \rho), \quad (4)$$

where

$$D(f, g) \equiv \frac{1}{2} \int \int f(x) g(y) |x - y|^{-1} dx dy, \quad (5)$$

for $1 < p < \infty$.

We shall be concerned with the following problem

$$\text{Min} \{ \xi_p(\rho) \mid \rho \in L^1 \cap L^p, \rho(x) \geq 0, \nabla \rho^{1/2} \in L^2 \text{ and } \int \rho(x) dx = \lambda \}, \quad (I)$$

where λ is a given positive constant, which, physically, is the total electron number.

Our main result is the following:

Theorem 1. *There is a critical value $0 < \lambda_c < \infty$ depending only on p and V such that*

- (a) If $\lambda \leq \lambda_c$, Problem (I) has a unique solution.
- (b) If $\lambda > \lambda_c$, Problem (I) has no solution. In addition,
- (c) When $p \geq \frac{4}{3}$, then $\lambda_c \geq Z \equiv \sum_{i=1}^k z_i$.
- (d) When $p \geq \frac{5}{3}$ and $k = 1$ (atomic case), then $\lambda_c > Z$.

Remark 1. Partial results were previously obtained by one of us. Namely in [4] it is proved that for the atomic case ($k = 1$) and $p = \frac{5}{3}$, Problem (I) has a solution if $\lambda \leq Z$.

Remark 2. Some of the open problems which are raised by our developments are the following:

- (i) Suppose $k \geq 2$ (molecular case) and $p = \frac{5}{3}$. Is $\lambda_c > Z$?
- (ii) Find estimates for λ_c .
- (iii) Is there binding for atoms?

With respect to the third problem, there is a non-rigorous argument of Balàzs [5] that indicates the possibility of binding for homopolar diatomic molecule in the TFW theory. Also, Gombás [6] applied TFW (including exchange corrections) to study the N_2 -molecule (i.e., $z_1 = z_2 = 7$) and found numerically that

there is binding. He actually computed the distance between the two centers to be $1.39\text{\AA}$ for the configuration of minimum energy. We do not give a proof of binding in the homopolar case, but we will prove that binding occurs for two very dissimilar atoms.

Remark 3. Theorem 1 obviously holds if we replace ξ_p by

$$\xi_p(\rho) = c_1 \int |\nabla \rho^{1/2}(x)|^2 dx + c_2 \int \rho^p(x) dx - \int V(x)\rho(x) dx + D(\rho, \rho),$$

where c_1 and c_2 are positive constants.

The proof of Theorem 1 is divided into several steps. In Sect. 1 we describe some basic properties of $\xi_p(\rho)$. In Sect. 2 we consider the problem

$$\text{Min} \{ \xi_p(\rho) \mid \rho \in D_p \}$$

where $D_p = \{ \rho \mid \rho(x) \geq 0, \rho \in L^3 \cap L^p, \nabla \rho^{1/2} \in L^2, D(\rho, \rho) < \infty \}$, and we prove that the minimum is achieved at a unique ρ_0 . Note that D_p contains $\{ \rho \mid \rho \geq 0, \rho \in L^3 \cap L^p \cap L^1, \nabla \rho^{1/2} \in L^2 \}$. We derive the Euler equation for ρ_0 . More precisely we set $\psi = \rho_0^{1/2}$ and we show that

$$-\Delta \psi + \psi^{2p-1} = \varphi \psi, \quad (6)$$

where

$$\varphi(x) = V(x) - \int \psi^2(y) |x - y|^{-1} dy. \quad (7)$$

In Sect. 3 we prove that $\psi \in L^2$ and we obtain some further properties of ψ (ψ is continuous, $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$, etc.). In Sect. 4 we show that if $p \geq \frac{4}{3}$ then $\int \psi^2(x) dx \geq Z$. In Sect. 5 we show that if $p \geq \frac{5}{3}$ and $k = 1$ then $\int \psi^2(x) dx > Z$. In addition $\psi(x) \leq M e^{-\delta|x|^{1/2}}$ for some appropriate constants M and $\delta > 0$. In Sect. 6 we prove that for every λ

$$E(\lambda) \equiv \text{Inf} \{ \xi_p(\rho) \mid \int \rho(x) dx = \lambda \} = \text{Inf} \{ \xi_p(\rho) \mid \int \rho(x) dx \leq \lambda \}$$

and we conclude the proof of Theorem 1. We also show that $E(\lambda)$ is convex, monotone non-increasing and that $E(\lambda)$ has a finite slope at $\lambda = 0$. This slope is the ground state energy of the corresponding one electron Schrödinger Equation. Using this last fact, binding for dissimilar atoms is proved.

1. Some Basic Properties of ξ_p

In Lemma 2, 3, 4 some properties which are useful in the study of Problem (I) are summarized.

Lemma 2. *For every $\varepsilon > 0$, there is a constant C_ε , depending on V but independent of ρ , such that*

$$\int V(x)\rho(x) dx \leq \varepsilon \|\rho\|_3 + C_\varepsilon D(\rho, \rho)^{1/2}, \quad (8)$$

for every $\rho \geq 0$.

Proof. Let $\delta > 0$ be a small constant and let $\zeta(x)$ be a smooth function such that

$0 \leq \zeta \leq 1$ and

$$\zeta(x) = \begin{cases} 1 & \text{on } \bigcup_{i=1}^k B_\delta(R_i) \\ 0 & \text{outside } \bigcup_{i=1}^k B_{2\delta}(R_i), \end{cases}$$

where $B_\delta(R_i)$ is the ball of radius δ and centered at R_i . δ is chosen such that all these ball $B_{2\delta}$ are disjoint. Let $V = V\zeta + V(1 - \zeta) \equiv V_1 + V_2$. Clearly $V_1 \in L^{3/2}$ and by choosing δ small enough we may assume that $\|V_1\|_{3/2} < \varepsilon$. Thus

$$\int V_1(x)\rho(x)dx \leq \varepsilon \|\rho\|_3. \quad (9)$$

On the other hand define the operator B to be

$$(B\rho)(x) = \int \rho(y) |x - y|^{-1} dy, \quad (10)$$

so that (in the sense of distributions) we have

$$-\Delta(B\rho) = 4\pi\rho.$$

Thus,

$$\int |\nabla(B\rho)|^2 dx = 8\pi D(\rho, \rho). \quad (11)$$

We deduce from (11) and Sobolev's inequality that

$$\|B\rho\|_6 \leq CD(\rho, \rho)^{1/2}. \quad (12)$$

Consequently,

$$\begin{aligned} \int V_2(x)\rho(x)dx &= \frac{1}{4\pi} \int (-\Delta V_2)(x)(B\rho)(x)dx \\ &\leq C \|\Delta V_2\|_{6/5} D(\rho, \rho)^{1/2} \end{aligned} \quad (13)$$

(note that $\Delta V_2 \in C_0^\infty$). Combining (9) and (13) we obtain the conclusion. $\square$

Lemma 3. *There exist positive constants α and C such that*

$$\xi_p(\rho) \geq \alpha(\|\rho\|_3 + \|\rho\|_p^p + \|\nabla\rho^{1/2}\|_2^2 + D(\rho, \rho)) - C$$

Proof. Use Lemma 2 and Sobolev's inequality. $\square$

Lemma 4. $\xi_p(\rho)$ is strictly convex.

Proof. The only non-standard fact is that the function $\rho \rightarrow \int |\nabla\rho^{1/2}|^2 dx$ is convex (or equivalently subadditive). Indeed let $\rho_1, \rho_2 \in D_p$ and set $\psi_1 = \rho_1^{1/2}$, $\psi_2 = \rho_2^{1/2}$, $\psi_3 = (\alpha\rho_1 + (1 - \alpha)\rho_2)^{1/2}$ with $0 < \alpha < 1$. Thus,

$$\begin{aligned} \psi_3 \nabla\psi_3 &= \alpha\psi_1 \nabla\psi_1 + (1 - \alpha)\psi_2 \nabla\psi_2 \\ &= (\alpha^{1/2}\psi_1)(\alpha^{1/2}\nabla\psi_1) + [(1 - \alpha)^{1/2}\psi_2][(1 - \alpha)^{1/2}\nabla\psi_2] \end{aligned}$$

and by Cauchy-Schwarz inequality

$$\psi_3 \nabla\psi_3 \leq (\alpha\psi_1^2 + (1 - \alpha)\psi_2^2)^{1/2} (\alpha|\nabla\psi_1|^2 + (1 - \alpha)|\nabla\psi_2|^2)^{1/2},$$

and therefore,

$$|\nabla\psi_3|^2 \leq \alpha |\nabla\psi_1|^2 + (1-\alpha) |\nabla\psi_2|^2. \quad \square$$

II. Minimization of $\xi_p(\rho)$ with $\rho \in D_p$ —The Euler Equation

We start with

Lemma 5. $\text{Min}\{\xi_p(\rho) \mid \rho \in D_p\}$ is achieved at some $\rho_0 \in D_p$.

Proof. Let $\rho_n \in D_p$ be a minimizing sequence. By Lemma 3 we have

$$\|\rho_n\|_3 \leq C, \|\rho_n\|_p \leq C, \|\nabla\rho_n^{1/2}\|_2 \leq C, D(\rho_n, \rho_n) \leq C.$$

Therefore, we may extract a subsequence, still denoted by ρ_n , such that

$$\rho_n \rightarrow \rho_0 \text{ weakly in } L^3 \text{ and in } L^p, \quad (14)$$

$$\rho_n \rightarrow \rho_0 \text{ a.e.}, \quad (15)$$

$$\nabla\rho_n^{1/2} \rightarrow \nabla\rho_0^{1/2} \text{ weakly in } L^2. \quad (16)$$

((15) relies on the fact that if Ω is a bounded smooth domain then $H^1(\Omega)$ is relatively compact in $L^2(\Omega)$. (14) and (16) implies that $\{\rho_n^{1/2}\}$ is bounded in $H^1(\Omega)$. Hence $\{\rho_n^{1/2}\}$ has a subsequence converging in $L^2(\Omega)$ and a.e.). Hence,

$$\begin{aligned} \liminf \int |\nabla\rho_n^{1/2}|^2 dx &\geq \int |\nabla\rho_0^{1/2}|^2 dx, \\ \liminf \int \rho_n^p(x) dx &\geq \int \rho_0^p(x) dx, \end{aligned}$$

$\liminf D(\rho_n, \rho_n) \geq D(\rho_0, \rho_0)$ (by Fatou's Lemma).

We now prove that

$$\int V(x)\rho_n(x) dx \rightarrow \int V(x)\rho_0(x) dx.$$

As in the proof of Lemma 2, we write $V = V_1 + V_2$. Clearly,

$$\int V_1(x)\rho_n(x) dx \rightarrow \int V_1(x)\rho_0(x) dx,$$

since $V_1 \in L^{3/2}$. On the other hand

$$\int V_2(x)\rho_n(x) dx = -\frac{1}{4\pi} \int (\Delta V_2)(B\rho_n) dx.$$

It follows from (12) that $B\rho_n \rightarrow B\rho_0$ weakly in L^6 . Thus,

$$\int V_2(x)\rho_n(x) dx \rightarrow \int V_2(x)\rho_0(x) dx.$$

Hence,

$$\xi_p(\rho_0) \leq \liminf \xi_p(\rho_n) = \text{Inf}\{\xi_p(\rho) \mid \rho \in D_p\}. \quad \square$$

We now derive the Euler Equation satisfied by ρ_0 . Set $\psi = \rho_0^{1/2}$.

Lemma 6. The minimizing ρ_0 satisfies

$$-\Delta\psi + \psi^{2p-1} = \varphi\psi, \quad (17)$$

where

$$\varphi = V - B\psi^2 \quad (18)$$

and (17) holds in the sense of distributions.

Proof. So far, we know that $\psi \in L^6 \cap L^{2p}$, $\nabla\psi \in L^2$ and $B\psi^2 \in L^6$. Since $V \in L^2_{\text{loc}}$, it follows that $\varphi \in L^2_{\text{loc}}$ and thus $\varphi\psi \in L^1_{\text{loc}}$. On the other hand $\psi^{2p-1} \in L^1_{\text{loc}}$ (since $\psi \in L^{2p}$). Therefore, (17) has a meaning in the sense of distributions. Consider the set $\tilde{D} = \{\zeta \in L^6 \cap L^{2p} \mid \nabla\zeta \in L^2 \text{ and } D(\zeta, \zeta) < \infty\}$. (Note that we do not assume $\zeta \geq 0$.) If $\zeta \in \tilde{D}$, then $\rho = \zeta^2 \in D_p$ and

$$\xi_p(\rho) = \int |\nabla\zeta|^2 dx + \frac{1}{p} \int \zeta^{2p} dx - \int V\zeta^2 + D(\zeta, \zeta) \equiv \phi(\zeta).$$

Indeed it suffices to recall that $\nabla\rho^{1/2} = \nabla|\zeta| = \nabla\zeta(\text{sgn}\zeta)$ (see [7]). Therefore, we find for every $\zeta \in \tilde{D}$

$$\phi(\psi) \leq \phi(\zeta)$$

Let $\eta \in C_0^\infty$; using the fact that $d/dt \phi(\psi + t\eta)|_{t=0} = 0$ we conclude easily that

$$\int \nabla\psi \cdot \nabla\eta dx + \int \psi^{2p-1}\eta dx = \int \varphi\psi\eta dx. \quad \square$$

III. Proof that the Minimizing $\psi \in L^2$ and Further Properties of ψ

We first prove that the minimizing ψ is continuous:

Lemma 7. ψ is continuous on $\mathbb{R}^3$; more precisely $\psi \in C^{0,\alpha}$ for every $\alpha < 1$ (i.e., for every bounded set $\Omega \subset \mathbb{R}^3$, there is a constant M such that $|\psi(x) - \psi(y)| \leq M|x - y|^\alpha$ $\forall x, y \in \Omega$).

Proof. We already know that $B\psi^2 \in L^6$ and (clearly) $V \in L^{3-\delta}_{\text{loc}} (\forall \delta > 0)$. Consequently, $\varphi \in L^{3-\delta}_{\text{loc}} (\forall \delta > 0)$. Since $\psi \in L^6$, it follows that $\varphi\psi \in L^{2-\delta}_{\text{loc}} (\forall \delta > 0)$. Therefore, we have

$$-\Delta\psi \leq f,$$

where $f = \varphi\psi \in L^{2-\delta}_{\text{loc}} (\forall \delta > 0)$ and in particular $f \in L^q_{\text{loc}}$ for some $q > 3/2$. We may, therefore, apply a result of Stampacchia (see [7], Théorème 5.2) to conclude that $\psi \in L^\infty_{\text{loc}}$. Going back to (17) and using the fact that $\psi \in L^\infty_{\text{loc}}$, we now see that $\Delta\psi \in L^{3-\delta}_{\text{loc}} (\forall \delta > 0)$. The standard elliptic regularity theory [8] implies that $\psi \in C^{0,\alpha}$ for every $\alpha < 1$. $\square$

We now prove an important property of ψ , namely, $\psi \in L^2$. Note that such a fact *cannot* be deduced from the knowledge that $\rho_0 \in D_p$. It is easy to construct a function $\rho \geq 0$ such that $\rho \in L^3 \cap L^p$, $\nabla\rho^{1/2} \in L^2$, $D(\rho, \rho) < \infty$ and $\int \rho(x) dx = \infty$.

Lemma 8. $\psi \in L^2$.

Proof. Suppose, by contradiction, that $\int \psi^2(x) dx = \infty$. Choose $r_1 > \max_{1 \leq i \leq k} |R_i|$ such that

$$\int_{|x| < r_1} \psi^2(x) dx \geq Z + 2\delta,$$

for some $\delta > 0$. We, thus, have

$$\begin{aligned} (B\psi^2)(x) &= \int \psi^2(y) |x - y|^{-1} dy \geq \int_{|y| < r_1} \psi^2(y) (|x| + |y|)^{-1} dy \\ &\geq (Z + 2\delta) / (|x| + r_1). \end{aligned}$$

Therefore,

$$\varphi(x) = V(x) - (B\psi^2)(x) \leq \frac{Z}{|x| - r_1} - \frac{Z + 2\delta}{|x| + r_1},$$

for $|x| > r_1$. Consequently, there is some $r_2 > r_1$ such that

$$\varphi(x) \leq -\delta |x|^{-1},$$

for $|x| > r_2$. It follows from (17) that

$$-\Delta\psi + \delta |x|^{-1} \psi \leq 0, \quad (19)$$

for $|x| > r_2$. We now use a comparison argument. Set

$$\tilde{\psi}(x) = M e^{-2(\delta|x|)^{1/2}},$$

where $M > 0$ is a constant. An easy computation shows that

$$-\Delta\tilde{\psi} + \delta |x|^{-1} \tilde{\psi} \geq 0, \quad (20)$$

for $x \neq 0$. Hence, by (19) and (20) we have

$$-\Delta(\psi - \tilde{\psi}) + \delta |x|^{-1} (\psi - \tilde{\psi}) \leq 0, \quad (21)$$

for $|x| > r_2$. We fix M in such a way that

$$\psi(x) \leq \tilde{\psi}(x), \quad (22)$$

for $|x| = r_2$ (this is possible since $\psi \in L_{\text{loc}}^\infty$). It follows from (21), (22) and the maximum principle that

$$\psi(x) \leq \tilde{\psi}(x), \quad (23)$$

for $|x| > r_2$. Since we only know that $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$ in a weak sense (namely $\psi \in L^6$), we must justify (23). We use a variant of Stampacchia's method. Fix $\zeta(x) \in C_0^\infty$ with $0 \leq \zeta \leq 1$ and

$$\zeta(x) = \begin{cases} 1 & \text{for } |x| < 1 \\ 0 & \text{for } |x| > 2. \end{cases}$$

Set $\zeta_n(x) = \zeta(x/n)$. Multiplying (21) by $\zeta_n(\psi - \tilde{\psi})^+$ (here we set $t^+ = \text{Max}(t, 0)$) and integrating on $[|x| > r_2]$, we find

$$\int_{|x| > r_2} \nabla(\psi - \tilde{\psi}) [\nabla \zeta_n (\psi - \tilde{\psi})^+ + \zeta_n \nabla(\psi - \tilde{\psi})^+] dx + \int_{|x| > r_2} \frac{\delta}{|x|} |(\psi - \tilde{\psi})^+|^2 \zeta_n dx \leq 0.$$

In particular it follows that

$$\int_{|x| > r_2} \frac{\delta}{|x|} |(\psi - \tilde{\psi})^+|^2 \zeta_n dx \leq \frac{1}{2} \int_{|x| > r_2} |(\psi - \tilde{\psi})^+|^2 \Delta \zeta_n dx. \quad (24)$$

But we have

$$\begin{aligned} \int_{|x| > r_2} |(\psi - \tilde{\psi})^+| \Delta \zeta_n dx &\leq \frac{C}{n^2} \int_{n < |x| < 2n} \psi^2(x) dx \\ &\leq C \left[\int_{n < |x| < 2n} \psi^6(x) dx \right]^{1/3}, \end{aligned}$$

by Hölder's inequality. Since $\psi \in L^6$, we conclude that the right side in (24) tends to zero as $n \rightarrow \infty$. Consequently,

$$\int_{|x| > r_2} \delta |x|^{-1} |(\psi - \tilde{\psi})^+|^2 dx = 0$$

and so $\psi \leq \tilde{\psi}$ for $|x| > r_2$. In particular $\int \psi^2(x) dx < \infty$, a contradiction with the initial assumption. $\square$

We now indicate some further properties of ψ .

Lemma 9. ψ is bounded on $\mathbb{R}^3$, $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$ and $\psi \in H^2$.

Proof. By (17) we have,

$$-\Delta \psi \leq V\psi, \quad (25)$$

and so

$$-\Delta \psi + \psi \leq (V + 1)\psi.$$

Clearly, $(V + 1)\psi \in L^2$ and so

$$\psi \leq (-\Delta + I)^{-1} [(V + 1)\psi]. \quad (26)$$

As is well known, the right side in (26) is bound and tends to zero as $|x| \rightarrow \infty$. Finally, note that $\psi^{2p-1} \leq C\psi$ for some constant C and $(B\psi^2)\psi \in L^2$ (since $\psi \in L^3$ and $B\psi^2 \in L^6$). Therefore, we conclude that $\Delta \psi \in L^2$ and so $\psi \in H^2$. $\square$

Lemma 10. $\psi > 0$ everywhere and ψ is C^∞ except at $x = R_i$ ($1 \leq i \leq k$).

Proof. From (17) we have,

$$-\Delta \psi + a\psi = 0,$$

where $a \in L^q_{\text{loc}}$ and $q > 3/2$. It follows from Harnack's inequality (see e.g. [9] Corollary 5.3) that either $\psi > 0$ everywhere or $\psi \equiv 0$. We now prove that $\psi \not\equiv 0$ by checking that $\text{Min}\{\xi_p(\rho) | \rho \in D_p\} < 0$. It clearly suffices to consider the case where $V(x) = z_1 |x|^{-1}$. Take the trial function $\rho^{1/2}(x) = \gamma \exp[-z_1 |x|/4]$. The terms in ρ which are homogeneous of degree one are $-\gamma^2 z_1^2/4$. The remaining terms are proportional to γ^s , $s > 2$. Hence for γ sufficiently small, $\xi_p(\rho) < 0$. Finally, the fact that ψ is C^∞ (except at R_i) follows easily from (17) by a standard bootstrap argument. $\square$

Remark. When $p \geq 3/2$, there is a simple estimate for ψ , namely

$$\psi^{2(p-1)}(x) \leq V(x), \quad (27)$$

for every x . Indeed set $u = \psi^{2(p-1)}(x) - V(x)$ so that for $x \neq R_i$ we have

$$\begin{aligned} \Delta u &= 2(p-1)\psi^{2p-3}(\Delta\psi) + (2p-2)(2p-3)\psi^{2p-4}|\nabla\psi|^2 \\ &\geq 2(p-1)\psi^{2p-3}(\Delta\psi) = 2(p-1)\psi^{2p-3}(\psi^{2p-1} - \varphi\psi). \end{aligned}$$

The function u achieves its maximum at some point $x_0 (\neq R_i)$. At x_0 we have $(\Delta u)(x_0) \leq 0$ and so $\psi^{2(p-1)}(x_0) \leq \varphi(x_0) \leq V(x_0)$. Thus $u(x_0) \leq 0$, and so $u(x) \leq 0$ everywhere.

IV. Proof that for the Minimizing ψ , $\int \psi^2(x) dx \geq Z$ (when $p \geq 4/3$)

We start with the following remark:

Lemma 11. For any $\zeta \in C_0^\infty$

$$-\int \frac{\Delta\psi}{\psi} \zeta^2 dx \leq \int |\nabla\zeta|^2 dx.$$

Proof. Integrate by parts and use the Cauchy-Schwarz inequality. □

We now prove the main estimate.

Lemma 12. When $p \geq 4/3$, $\int \psi^2(x) dx \geq Z$.

Proof. Let $\zeta_0 \in C_0^\infty$ be a spherically symmetric function such that $\zeta_0 \neq 0$, $\zeta_0(x) = 0$ for $|x| < 1$ and for $|x| > 2$. Set $\zeta_n(x) = \zeta_0(x/n)$. By (17) we have,

$$-\int \frac{\Delta\psi}{\psi} \zeta_n^2 dx + \int \psi^{2p-2} \zeta_n^2 dx = \int \varphi \zeta_n^2 dx. \quad (28)$$

Using Lemma 11, we find

$$-\int \frac{\Delta\psi}{\psi} \zeta_n^2 dx \leq \int |\nabla\zeta_n|^2 dx \leq Cn. \quad (29)$$

Next we claim that, if $p \geq 4/3$, then

$$\int \psi^{2p-2} \zeta_n^2 dx \leq \varepsilon_n n^2, \quad (30)$$

where $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Indeed we have

$$\int \psi^{2p-2} \zeta_n^2 dx \leq \int_{n < |x| < 2n} \psi^{2p-2} dx.$$

If $2p-2 \geq 2$, we use the fact that $\psi^{2p-2} \leq C\psi^2$ in order to obtain (30). If $2p-2 \leq 2$, we use Hölder's inequality and we find

$$\int_{n < |x| < 2n} \psi^{2p-2} dx \leq C \left[\int_{n < |x| < 2n} \psi^2 dx \right]^{p-1} (n^3)^{2-p}.$$

Assuming $p \geq 4/3$, we deduce (30). On the other hand, since ζ_n is spherically symmetric, we have

$$\int \varphi \zeta_n^2 dx = \int [\varphi] \zeta_n^2 dx,$$

where $[\varphi]$ denotes the spherical average of φ , i.e.,

$$[\varphi](x) = \frac{1}{4\pi|x|^2} \int_{|\Omega|=|x|} \varphi(\Omega) d\Omega = \frac{1}{4\pi} \int_{|\Omega|=1} \varphi(|x|\Omega) d\Omega.$$

By a result of [1] (Eq. (35)) we know that

$$[\varphi](x) \geq (Z - \lambda_0)/|x|, \text{ for } |x| > \max_{1 \leq i \leq k} |R_i|, \quad (31)$$

where $\lambda_0 = \int \psi^2(x) dx$. Hence, for large n , we find

$$\int \varphi \zeta_n^2 dx \geq (Z - \lambda_0) \int \frac{1}{|x|} \zeta_n^2(x) dx = \alpha(Z - \lambda_0)n^2, \quad (32)$$

for some positive constant α . Combining (28), (29), (30), and (32), we find

$$\alpha(Z - \lambda_0)n^2 \leq Cn + \varepsilon_n n^2.$$

As $n \rightarrow \infty$, we conclude that $Z \leq \lambda_0$. □

Remark. Lemma 12 can also be proved by a direct variational calculation using (4). This is given in Theorem 4.10 of reference [4].

V. $\int \psi^2(x) dx > Z$ when $p \geq 5/3$ and $k = 1$ (Atomic Case)

We assume now that $V(x) = Z|x|^{-1}$. The main result is the following:

Lemma 13. ([4], Theorem 4.13) Assume $p \geq 5/3$, then $\int \psi^2(x) dx > Z$. In addition $\psi(x) \leq M e^{-2(\delta|x|)^{1/2}}$ for some constants M and $0 < 2\delta < \int \psi^2(x) dx - Z$.

Proof. Since the solution of the problem $\text{Min} \{ \xi_p(\rho) | \rho \in D_p \}$ is unique, it follows that ρ_0 —and therefore ψ —is spherically symmetric. In particular $\varphi = V - B\psi^2$ is also spherically symmetric. On the other hand by (31) we have,

$$\varphi(x) = [\varphi](x) \geq (Z - \lambda_0)/|x|, \quad (33)$$

for $x \neq 0$. We already know that $\int \psi^2(x) dx \geq Z$; suppose by contradiction that $\int \psi^2(x) dx = Z$. By (33) we have $\varphi \geq 0$ and consequently (from (17))

$$-\Delta\psi + \psi^{2p-1} \geq 0, \quad (34)$$

for $x \neq 0$. We now use a comparison function. Set $\tilde{\psi}(x) = C|x|^{-3/2}$. An easy computation shows that

$$-\Delta\tilde{\psi} + \tilde{\psi}^{2p-1} \leq 0, \text{ for } |x| > 1, \quad (35)$$

provided $0 < C \leq C_0$ where $C_0^{2p-2} = 3/4$. We fix the constant $0 < C \leq C_0$ such that

$$\tilde{\psi}(x) \leq \psi(x) \text{ for } |x| = 1, \quad (36)$$

(Recall that by Lemma 10, $\psi > 0$). It follows from (34), (35), (36) and the (usual) maximum principle that $\tilde{\psi}(x) \leq \psi(x)$ for $|x| > 1$. Since $\tilde{\psi} \notin L^2(|x| > 1)$, we obtain a contradiction. Therefore, $\int \psi^2(x) dx > Z$; finally we argue as in the proof of

Lemma 8 and we conclude that for some M , $\psi(x) \leq M \exp(-2(\delta|x|)^{1/2})$ where $2\delta < \int \psi^2(x)dx - Z$. $\square$

Remark. If the assumptions of Lemma 13 are not satisfied, it may happen that $\int \psi^2 dx = Z$. Consider for example the functional

$$\xi(\rho) = c \int |\nabla \rho^{1/2}(x)|^2 dx + 2/3 \int \rho^{3/2}(x) dx - \int V(x)\rho(x) dx + D(\rho, \rho),$$

where $V(x)$ is given by (3). We claim that if $c \leq 1/16\pi$, then we have $\int \psi^2(x)dx = Z$. Indeed set $u = \psi - 2\varphi$. Recall that $-c\Delta\psi + \psi^2 = \varphi\psi$, and

$$\Delta\varphi = 4\pi\psi^2 \text{ if } x \neq R_i.$$

Thus,

$$\Delta u = \frac{1}{c}(\psi^2 - \varphi\psi) - 8\pi\psi^2 \geq \frac{1}{c}(\psi^2 - \varphi\psi) - \frac{1}{2c}\psi^2 = \frac{\psi u}{2c}.$$

At a point x_0 where u achieves its maximum we have $(\Delta u)(x_0) \leq 0$ and so $u(x_0) \leq 0$ (note that $u(x) \rightarrow 0$ as $|x| \rightarrow \infty$). Consequently, $u \leq 0$ everywhere and so $\varphi \geq 0$ everywhere. Therefore, we must have $Z \geq \int \psi^2(x)dx$. Since we already know that $Z \leq \int \psi^2(x)dx$, it follows that $\int \psi^2(x)dx = Z$.

If $c > 1/16\pi$, one can still prove the weaker inequality

$$c\psi \leq \frac{1}{8\pi}\varphi + (c - 1/16\pi)V, \quad (37)$$

by the same type of arguments. Using (37) we have

$$Z \leq \int \psi^2 dx \leq Z(1 + 8\pi(c - 1/16\pi)),$$

because $\varphi + \alpha V \geq 0$ (with $\alpha > 0$) implies $\int \psi^2(x)dx \leq Z(1 + \alpha)$.

Note that in the one center (atomic) case $\xi_{3/2}$ is scale invariant. In fact, $E_{3/2} = \text{Min } \xi_{3/2} = cZ^3$ and $\psi(x) = Z^2\psi(Zx)$.

VI. Proof of Theorem 1 Concluded

We need a final Lemma.

Lemma 14. ([4], Theorem 4.2) For every $\lambda > 0$ we have $\text{Inf}\{\xi_p(\rho) | \rho \in D_p \text{ and } \int \rho(x)dx = \lambda\} = \text{Inf}\{\xi_p(\rho) | \rho \in D_p \text{ and } \int \rho(x)dx \leq \lambda\}$.

Proof. Let $\rho \in D_p$ be such that $\int \rho(x)dx < \lambda$; the Lemma is an obvious consequence of the following claim: There exists a sequence $\rho_n \in D_p$ such that $\int \rho_n(x)dx = \lambda$ and $\liminf \xi_p(\rho_n) \leq \xi_p(\rho)$. As ρ_n , we choose

$$\rho_n(x) = \rho(x) + \frac{k}{n^3} \zeta_n^2(x)$$

where $\zeta_n(x) = \zeta_0(x/n)$ ($\zeta_0 \in C_0^\infty$ is any function $\zeta_0 \neq 0$) and $k = [\lambda - \int \rho(x)dx] / \int \zeta_0^2(x)dx$, so that $\int \rho_n(x)dx = \lambda$. We now check that $\liminf \xi_p(\rho_n) \leq \xi_p(\rho)$. Using the subadditivity of the function $\int |\nabla \rho^{1/2}(x)|^2 dx$ and the convexity of ρ^p , we find

$$\xi_p(\rho_n) \leq \xi_p(\rho) + A_n + B_n + C_n,$$

where

$$\begin{aligned} A_n &= \frac{k}{n^3} \int |\nabla \zeta_n(x)|^2 dx \\ B_n &= \frac{k}{n^3} \int \rho_n^{p-1} \zeta_n^2 dx \\ C_n &= \frac{k}{n^3} \int (B\rho) \zeta_n^2 dx + \frac{k^2}{2n^6} \int (B\zeta_n^2) \zeta_n^2 dx \end{aligned}$$

We shall prove that A_n, B_n, C_n tend to zero. Indeed we have

$$A_n = \frac{k}{n^2} \int |\nabla \zeta_0(x)|^2 dx \rightarrow 0.$$

Next, by Hölder's inequality

$$B_n \leq \frac{k}{n^3} \left[\int \rho_n^p(x) dx \right]^{p-1/p} \left[\int \zeta_n^{2p}(x) dx \right]^{1/p}.$$

But,

$$\int \rho_n^p(x) dx \leq C \int [\rho^p(x) + ((k/n^3)\zeta_n^2(x))^p] dx \leq C$$

and so $B_n \rightarrow 0$. Finally, we have

$$C_n \leq \frac{k}{n^3} \left[\int (B\rho) \rho dx \right]^{1/2} \left[\int (B\zeta_n^2) \zeta_n^2 dx \right]^{1/2} + \frac{k^2}{2n^6} \int (B\zeta_n^2) \zeta_n^2 dx$$

and

$$\frac{1}{n^6} \int (B\zeta_n^2) \zeta_n^2 = \frac{2}{n} D(\zeta_0^2, \zeta_0^2).$$

Therefore, $C_n \rightarrow 0$. □

Proof of Theorem 1 concluded: For every $\lambda > 0$ we set $E(\lambda) = \inf\{\xi_p(\rho) \mid \rho \in D_p \text{ and } \int \rho(x) dx \leq \lambda\}$. It is clear that $E(\lambda)$ is non-increasing and that $E(\lambda)$ is convex. In addition, the same proof as in Lemma 5 shows that there is a unique $\rho_\lambda \in D_p$ such that $\int \rho_\lambda(x) \leq \lambda$ and $\xi_p(\rho_\lambda) = E(\lambda)$. Set $\lambda_c = \int \psi^2(x) dx$. It is clear that for $\lambda \geq \lambda_c$ the function $E(\lambda)$ is constant: $E(\lambda) = E(\lambda_c)$; while $E(\lambda)$ is strictly decreasing on the interval $[0, \lambda_c]$. It follows that for $\lambda \leq \lambda_c$ we have $\int \rho_\lambda(x) dx = \lambda$. Consequently, if $\lambda \leq \lambda_c$ there is a unique solution for Problem (I). When $\lambda > \lambda_c$, we deduce from Lemma 14 that Problem (I) has no solution. □

By the same methods as used in Lemma 6, the unique minimizing ψ for $E(\lambda)$, $\lambda \leq \lambda_c$ satisfies

$$-\Delta\psi + U_\psi\psi = -\phi_0\psi, \quad (38)$$

where,

$$U_\psi = \psi^{2p-2} - V + B\psi^2. \quad (39)$$

It is also true (as shown in [1] for the TF problem) that $E(\lambda)$ is differentiable and

$$-\phi_0(\lambda) = \partial E / \partial \lambda. \quad (40)$$

Since ψ satisfies (38) and $\psi(x) > 0$ (by Harnack's inequality as in Lemma 10), we can conclude that ψ and ϕ_0 are, in fact, the lowest eigenfunction and eigenvalue of $-\Delta + U_\psi(x)$.

To summarize what has been proved so far, the function $E(\lambda)$ has many features in common with the $E(\lambda)$ for TF theory [1]. It is convex, non-increasing and has an absolute minimum at some $\lambda_c < \infty$ beyond which $E(\lambda)$ is constant. One important difference is that $\lambda_c > Z$, at least for the atomic case and $p \geq 5/3$. There is another important difference: in TF theory, $E(\lambda) \sim -c\lambda^{1/3}$ for small λ , i.e., $\partial E/\partial \lambda|_{\lambda=0} = -\infty$. In TFW theory, this is not so as the following shows.

Lemma 15. *Let e_0 be the lowest eigenvalue (ground state energy) of the Schrödinger operator $-\Delta - V(x)$ with V given by (3). Then*

$$\lim_{\lambda \downarrow 0} \lambda^{-1} E(\lambda)/e_0 = 1, \text{ i.e.,} \\ -\phi_0(0) = \partial E(\lambda)/\partial \lambda|_{\lambda=0} = e_0.$$

Proof. Let φ be the normalized eigenfunction of $-\Delta - V(x)$ belonging to e_0 and let $\rho_\lambda(x) = \lambda \varphi(x)^2$. Then $\xi_p(\rho_\lambda) \leq \lambda e_0 + 0(\lambda)$ since the terms in ξ_p of degree higher than the first, while they are positive, are finite and $0(\lambda)$. Conversely,

$$E(\lambda) \geq \inf_{\rho} \left\{ \int |\nabla \rho^{1/2}|^2 - \int V \rho \mid \int \rho = \lambda \right\},$$

but this is precisely the variational problem for the Schrödinger Equation. $\square$

Yet another important distinction with TF theory should be noted. $\psi(x)$ is never zero and therefore $\rho(x) = \psi(x)^2$ does not have compact support. In TF theory, ρ has compact support [1] whenever $\lambda < \lambda_c$.

Binding of Atoms in TFW Theory

Binding does not occur in TF theory [1]; that is Teller's Theorem. Binding can occur in TFW theory as we shall now prove in a special case.

First it is necessary to have a clear definition of what binding means. Given the nuclear coordinates $R_1, \dots, R_k$, define

$$\tilde{E}(\lambda; \{R_i\}) \equiv E(\lambda; \{R_i\}) + U(\{R_i\}) \quad (41)$$

where

$$U(\{R_i\}) = \sum_{1 \leq i < j \leq k} z_i z_j |R_i - R_j|^{-1}. \quad (42)$$

Let us start with two neutral atoms infinitely far apart. The total energy is then $A_{12} = E_1(z_1) + E_2(z_2)$ where E_i is the energy of an atom with nuclear charge z_i . The next step is to distribute the total electron charge so as to minimize the total energy, namely

$$B_{12} = \min_{0 \leq \lambda \leq z_1 + z_2} (E_1(z_1 + z_2 - \lambda) + E_2(\lambda)). \quad (43)$$

It may be, and usually is the case in TFW theory, that $B_{12} < A_{12}$, i.e., ions are more stable than atoms. Finally, we bring the atoms together and define

$$C_{12} = \inf_{R_1, R_2} \tilde{E}(z_1 + z_2; R_1, R_2). \quad (44)$$

If $C_{12} < B_{12}$, then binding occurs.

To show that binding is possible, suppose that z_2 is sufficiently small compared to z_1 so that the following is satisfied

$$(a) \quad (\partial E_1 / \partial \lambda) \left(\frac{z_1 + \lambda_c^1}{2} \right) < -z_2^2/4,$$

$$(b) \quad \lambda_c^1 - z_1 \geq 2z_2,$$

where λ_c^1 is the critical λ for atom 1; then we claim that

$$B_{12} = E_1(z_1 + z_2). \quad (45)$$

This follows from the following observation: $(\partial E_2 / \partial \lambda) \geq (\partial E_2 / \partial \lambda)(0) = -z_2^2/4$ by Lemma 15 and convexity. Then the equation $(\partial E_1 / \partial \lambda)(z_1 + z_2 - \lambda) = (\partial E_2 / \partial \lambda)(\lambda)$ cannot have a solution for $0 \leq \lambda \leq z_1 + z_2$.

For two such atoms the lowest total energy occurs when the smaller atom is completely stripped of its electrons which become attached to the larger atom. Now consider the first atom with $R_1 = 0$ and with $\lambda = z_1 + z_2 \leq \lambda_c^1$. The electric potential φ which is spherically symmetric, will be negative for large R . If the second nucleus z_2 is placed at a point r where $\varphi(r) < 0$, the total energy will be reduced by an amount $z_2 \varphi(r)$. Thus,

$$C_{12} < B_{12} + z_2 \varphi(r)$$

and binding occurs.

Acknowledgements. One of the authors (H.B.) would like to thank the following for their hospitality during the course of this work: The Princeton University, Courant Institute and The University of Chicago.

References

1. Lieb, E. H., Simon, B.: Adv. Math. **23**, 22–116 (1977)
2. von Weizsäcker, C. F.: Z. Phys. **96**, 431–458 (1935)
3. Kompaneets, A. S., Pavloskii, E. S.: Sov. Phys. JETP **4**, 328–336 (1957)
4. Benguria, R.: The von Weizsäcker and exchange corrections in the Thomas–Fermi theory. Princeton University Thesis: June 1979 (unpublished)
5. Balázs, N. L.: Phys. Rev. **156**, 42–47 (1967)
6. Gombás, P.: Acta Phys. Hung. **9**, 461–469 (1959)
7. Stampacchia, G.: Equations elliptiques du second ordre à coefficients discontinus. Montreal: Presses de l'Univ. 1965
8. Bers, L., Schechter, M.: Elliptic equations in Partial Differential Equations. New York: Interscience pp. 131–299. 1964
9. Trudinger, N.: Ann. Scuola Norm. Sup. Pisa **27**, 265–308 (1973)

Communicated by A. Jaffe

Received June 9, 1980

Analysis of the Thomas-Fermi-von Weizsäcker Equation for an Infinite Atom Without Electron Repulsion

Elliott H. Lieb*

Departments of Mathematics and Physics, Princeton University, P.O.B. 708, Princeton, NJ 08544, USA

Abstract. The equation

$$\{-\Delta + |\psi(x)|^{2p-2} - |x|^{-1}\} \psi(x) = 0$$

in three dimensions is investigated. Uniqueness and other properties of the positive solution are proved for $3/2 < p < 2$. There are two physical interpretations of this equation for $p = 5/3$: (i) As the TFW equation for an infinite atom *without* electron repulsion; (ii) The positive solution, ψ , suitably scaled, is asymptotically equal to the solution of the TFW equation for an atom or molecule *with* electron repulsion in the regime where the nuclear charges are large and x is close to one of the nuclei.

I. Introduction

The equation to be analyzed here of primary physical interest is

$$\{-\Delta + |\psi(x)|^{4/3} - |x|^{-1}\} \psi(x) = 0 \quad (1.1)$$

in three dimensions and with ψ real valued. (1.1) was introduced in [9], wherein it was asserted without proof that (1.1) has a unique, positive solution. The present paper contains that proof. If $z, \gamma, A > 0$ and

$$\tilde{\psi}(z, \gamma, A, x) \equiv (z^2/A\gamma)^{3/4} \psi(zx/A), \quad (1.2)$$

then

$$\{-A\Delta + \gamma|\tilde{\psi}(x)|^{4/3} - z|x|^{-1}\} \tilde{\psi}(x) = 0. \quad (1.3)$$

and conversely. Thus, (1.3) and (1.1) are equivalent problems.

(1.3) is to be compared with the Thomas-Fermi-von Weizsäcker (TFW) equation for a molecule [2-4, 9, 10, 13, 14]:

$$\{-A\Delta + \gamma|\hat{\psi}(x)|^{4/3} - V(x) + (|x|^{-1} * \hat{\psi}^2)(x)\} \hat{\psi}(x) = -\mu\hat{\psi}(x), \quad (1.4)$$

* Work partially supported by U.S. National Science Foundation grant PHY-7825390 A02

with $V(x) = \sum_{j=1}^k z_j |x - R_j|^{-1}$ being the potential of k nuclei of charges $z_j \geq 0$ located at R_j . The term $|x|^{-1} * \hat{\psi}^2$ is called “the electron repulsion.” (1.3) has a unique positive solution (denoted by $\hat{\psi}$) for all $\mu \geq 0$ [2, 3, 9]. $-\mu$ is the “chemical potential.” For all $\mu \geq 0$, $\psi \in L^2(\mathbb{R}^3)$ [3, 9].

(1.3) is seen to be the TFW equation for an atom ($k=1$, $z_1=z$, $R_1=0$), but with the electron repulsion omitted. It will be shown here that (1.1), which is equivalent to (1.3), also has a unique positive solution, ψ , and $F(\psi) < \infty$ [cf. (1.8)]. The fact that $\mu=0$ in (1.1) and (1.3) means that the electron number, $\lambda = \int \hat{\psi}^2$, is maximal. We shall see that in this case [for (1.1) and (1.3) but not (1.4)], $\lambda = \infty$, which is physically reasonable since the maximum electron number for the quantum mechanical Bohr atom (without repulsion) is also infinity.

If the foregoing were the only interpretation of (1.3) it would not be especially interesting. However, (1.3) has another interpretation as proved in [9]: Consider the full TFW equation, (1.4), with the scaling $z_j = az_j^0$, $R_j = a^{-1/3} R_j^0$ and $\lambda = a\lambda^0$ with $\lambda^0 > 0$. If $a \rightarrow \infty$ (with A, γ fixed), and if x is close to one of the R_j then $\hat{\psi}(x) \rightarrow \tilde{\psi}(x - R_j)$ (with $z \equiv z_j$) in the following sense: Fix $y \in \mathbb{R}^3$ then

$$\lim_{a \rightarrow \infty} z_j^{-3/2} \hat{\psi}(R_j + z_j^{-1}y) = \lim_{a \rightarrow \infty} z_j^{-3/2} \tilde{\psi}(z_j^{-1}y) = (A\gamma)^{-3/4} \psi(y/A), \quad (1.5)$$

where ψ is the positive solution of (1.1) and with the convergence being pointwise and in L^1_{loc} . [Note that the second expression in (1.5) is, in fact, independent of a . The right side of (1.5) is independent of λ .]

Not only does $\hat{\psi} \rightarrow \tilde{\psi}$ as in (1.5) but also the difference between the TFW energy and the TF energy is given, to leading order in a , by ψ :

$$E^{\text{TFW}} - E^{\text{TF}} = D \sum_{j=1}^k z_j^2 + o(a^2), \quad (1.6)$$

where the constant D is given by [9]

$$D = A^{1/2} \gamma^{-3/2} F(\psi), \quad (1.7)$$

and where the functional F is defined generally for all real ψ by

$$F(\psi) = \int |\nabla \psi|^2 + \int k(\psi(x), x) dx \quad (1.8)$$

with

$$k(\psi, x) \equiv \frac{3}{5} |\psi|^{10/3} - |x|^{-1} \psi^2 + \frac{2}{5} |x|^{-5/2} \geq 0. \quad (1.9)$$

Note that $k(\psi, x) > 0$ and $k(\psi, x) = 0$ if and only if $\psi = |x|^{-3/4}$.

The positive solution of (1.1) has been evaluated numerically [8] with the result that

$$\psi(0) = 0.9701330$$

$$F(\psi) = 17.1676.$$

Generalization. Equation (1.1) has the following obvious generalization

$$\Delta \psi(x) = \{|\psi(x)|^{2p-2} - |x|^{-1}\} \psi(x) \quad (1.10)$$

with $p > 1$. The physical case, (1.1), corresponds to $p = 5/3$. The TFW equation, (1.4) can be generalized in the same way; this was in fact done in [2, 3, 9]. In TF theory, $p = 4/3$ and $3/2$ play a special role, while $p = 4/3$, $3/2$, and $5/3$ are special in TFW theory [9]. In the analysis here of (1.10), $p = 3/2$ and 2 are special (Theorems 3–5 and 7–9). Fortunately, the physical value $5/3$ is in the range $3/2 < p < 2$ in which all theorems are applicable.

The appropriate energy functional is (1.8), but with $k(\psi, x)$ replaced by

$$k_p(\psi, x) = \frac{1}{p} |\psi|^{2p} - |x|^{-1} \psi^2 + \frac{p-1}{p} |x|^{-p/(p-1)} \geq 0. \quad (1.11)$$

The corresponding F will be denoted by F_p . However, as will be seen in Sect. III, F_p is useful only when $3/2 < p < 2$. It is for this range of p that the existence of a positive solution will be proved.

One of the more amusing technical exercises is in Sect. IV where an asymptotic expansion for ψ is established. While the expansion is heuristically obvious, its proof is not, primarily because Δ in (1.10) is essentially a singular perturbation.

II. Properties of Eq. (1.10)

Initially, (1.10) will be interpreted as a distributional equation. Although our main interest is in positive solutions, we shall not restrict ourselves to such and will assume only that ψ is a real valued function. It will turn out that the class of functions such that $F_p(\psi) < \infty$ is the natural class to consider, when $3/2 < p < 2$, but this will not be assumed initially. The only assumption to be understood in all the theorems is that $p > 1$ and

$$\psi \in M = \{\psi \mid \nabla \psi \in L^2_{\text{loc}}, \psi \in L^1_{\text{loc}}\}. \quad (2.1)$$

We begin with two “local” theorems.

Theorem 1. *Let B be a bounded open ball in $\mathbb{R}^3$. Let ψ satisfy (1.10) in B in the sense of distributions with $\psi \in L^1(B)$ and $\nabla \psi \in L^2(B)$. Then*

- (i) *ψ is continuous. More precisely, $\psi \in C^{0,\alpha}(\bar{B})$ for all $0 < \alpha < 1$. I.e. there is a constant, C , such that $|\psi(x) - \psi(y)| \leq C|x - y|^\alpha$ for all $x, y \in B$.*
- (ii) *If $\psi(x) \geq 0$, $\forall x \in B$, then either $\psi(x) \equiv 0$ or else $\psi(x) > 0$, $\forall x \in B$.*
- (iii) *If $0 \notin B$ then $\psi \in C^{1,1/2}(\bar{B})$.*
- (iv) *If $0 \notin B$ and $\psi(x) \geq 0$, $\forall x \in B$, then ψ is real analytic in B .*

Proof. (i) $\psi \in W^{1,1}$. By the Sobolev imbedding theorem [1],

$$\begin{aligned} \psi \in W^{0,3/2} = L^{3/2} &\Rightarrow \psi \in W^{1,3/2} \Rightarrow \psi \in L^3 \Rightarrow \psi \in W^{1,2} \Rightarrow \psi \in L^6 \\ &\Rightarrow f \equiv |x|^{-1} \psi \in L^{2-\varepsilon}, \forall \varepsilon > 0; \end{aligned}$$

in particular $f \in L^q$ for some $q > 3/2$.

If $p < 5/2$ then $\psi \in L^6 \Rightarrow |\psi|^{2p-1} \in L^{3/2+\varepsilon}$ for some $\varepsilon > 0$. Then $\psi \in W^{2,\varepsilon+3/2}$ and thus $\psi \in C^{0,\alpha}$ for some $\alpha > 0$. Then the right side of (1.10) is in $L^{3-\varepsilon}$, all $\varepsilon > 0$, so $\psi \in W^{2,3-\varepsilon}$. Then $\psi \in C^{0,\alpha}$ for all $\alpha < 1$. This proves (i) for $p < 5/2$.

If $p \geq 5/2$ a different argument is needed. The fact that (1.10) holds as a distributional equation means that the right side [call it $h(x)$] must be a distribution and hence must be in L^1_{loc} . Since $|x|^{-1}\psi \in L^1$, $|\psi|^{2p-1} \in L^1_{\text{loc}}$. Kato's inequality [7], $\Delta|\psi| \geq (\text{sgn } \psi)\Delta\psi$ as distributions (with $\text{sgn } \psi = |\psi|/\psi$ if $\psi \neq 0$ and $\text{sgn } \psi = 0$ if $\psi = 0$), then implies that

$$\Delta|\psi| \geq (\text{sgn } \psi) \{|\psi|^{2p-2} - |x|^{-1}\} \psi = |\psi|^{2p-1} - |x|^{-1}|\psi|. \quad (2.2)$$

Therefore $-\Delta|\psi| \leq |x|^{-1}|\psi| \equiv f$. A result of Stampacchia [12, Théorème 5.2] is that if $f \in L^q_{\text{loc}}$ with $q > 3/2$ then $|\psi| \in L^\infty_{\text{loc}}$. But our f satisfies this condition. Returning to (1.10), $h \in L^{3-\varepsilon}_{\text{loc}}$, $\varepsilon > 0$, and (i) is thus proved as before.

(ii) We have $\Delta\psi = g\psi$ with $g \in L^q$ and $q > 3/2$. By the Harnack inequality (cf. [5]), $\psi \equiv 0$ or $\psi(x) > 0$, $\forall x$.

(iii) $\psi, \Delta\psi \in L^\infty$ and $\nabla\psi \in L^2 \Rightarrow \psi \in W^{2,2} \Rightarrow \nabla\psi \in L^6 \Rightarrow \psi \in W^{2,6} \Rightarrow \psi \in C^{1,1/2}$ (see [1]).

(iv) Assume $\psi(x) > 0$, whence $|\psi|^{2p-1}$ has as many derivatives as ψ has. By a bootstrap argument ψ is C^∞ (see [10, Theorem IV.5]). Since $|x|^{-1}$ is real analytic for $x \neq 0$, by [11, Theorem 5.8.6] ψ is real analytic. $\square$

Theorem 2 (cusp condition [6]). Assume the hypotheses of Theorem 1 and also that $0 \in B$. Then

$$\psi(0) = -2 \lim_{r \downarrow 0} \int \Omega \cdot \nabla \psi(r\Omega) d\Omega, \quad (2.3)$$

where $d\Omega$ is the normalized invariant measure on the unit sphere. In particular, if ψ is spherically symmetric about zero, then

$$\psi(0) = -2 \lim_{r \downarrow 0} d\psi(r)/dr. \quad (2.4)$$

Proof. Simply integrate (1.10) by parts. $\square$

Theorem 3. Assume that $p \geq 3/2$. Suppose ψ satisfies (1.10) in the sense of distributions on all of $\mathbb{R}^3$. Then $|\psi(x)| < |x|^{-1/(2p-2)}$.

Remark. Some condition on p is needed and we believe $p \geq 3/2$ is the right one. If $p < 3/2$ there cannot be any positive ψ satisfying both (1.10) and $\psi(x) \leq |x|^{-1/(2p-2)}$. For then $0 \leq h = -\Delta\psi \leq |x|^{(1-2p)/(2p-2)} \equiv f(x)$, where $-h$ is the right side of (1.10). Since $g \equiv |x|^{-1} * f$ is finite for $|x| > 0$ and $g(x) \rightarrow 0$ as $|x| \rightarrow \infty$, $\psi = 4\pi|x|^{-1} * h$. Since $h \neq 0$, this implies that $\psi(x) \geq c|x|^{-1}$ for $|x| > \text{some } R$ and $c > 0$. This is a contradiction. Thus, $p \geq 3/2$ is the right condition for positive ψ . It is possible that (1.10) has no positive solution even when $p = 3/2$, for in that case $\psi(x) = |x|^{-1}$ satisfies (1.10) everywhere except at the origin.

Proof. By Theorem 1 we can assume that ψ is continuous. Let $b = -1/(2p-2) \geq -1$ (it is here that $p \geq 3/2$ enters). $\Delta|x|^b \leq 0$. If $g(x) \equiv |\psi(x)| - |x|^b$ we have, by (2.2), that $\Delta g > 0$ on the set $A = \{x | g(x) > 0\}$. Since ψ is continuous (Theorem 1) we have that (i) A is open; (ii) $0 \notin A$; (iii) $h(x) \equiv \max[g(x), 0]$ is subharmonic on $\mathbb{R}^3$, i.e. $\Delta h \geq 0$. If we knew that $h(x) \rightarrow 0$ as $|x| \rightarrow \infty$ we could then conclude that $h \equiv 0$ and thus that $|\psi(x)| \leq |x|^b$. But this has to be proved.

We have, in fact, that on A

$$\Delta h \geq \{|\psi|^{-1/b} - |x|^{-1}\} |\psi|. \quad (2.5)$$

Since $|\psi| = h + |x|^b$ on A , and $(\alpha + \beta)^t > \alpha^t + \beta^t$ for $\alpha, \beta \geq 0$, $t \geq 1$,

$$\Delta h \geq h^{2p-1} \quad \text{on } \mathbb{R}^3. \quad (2.6)$$

Now let $f(x) = f(|x|)$ be the spherical average of h , i.e. $f(r) = \int h(\Omega r) d\Omega$ with $d\Omega$ being the normalized invariant measure on the unit sphere. By averaging (2.6), and using $\text{av}\{h^t\} \geq \{\text{av}h\}^t$ for $t \geq 1$, we have that $\Delta f \geq f^{2p-1}$. With $r = |x|$, let $rf(r) \equiv u(r)$, whence

$$u'' \geq r^{-2p+2} u^{2p-1} \quad \text{for } r > 0. \quad (2.7)$$

Assume that $u(r) \not\equiv 0$. Since $0 \notin A$, there is a $R_0 > 0$ such that $u(r) = 0$ for $0 \leq r < R_0$. Since $u'' \geq 0$ and $u \geq 0$, u is continuous, convex and non-decreasing in r . Therefore for some $R_1 > 0$, $u(R_1) > 0$ and $u'(R_1) > 0$. Thus $u(r) > br$ for $r > \text{some } R_2$ and with $b > 0$.

Let $w_R(x) \equiv \sigma R(R-r)^{2b}$ be defined on $D_R = \{r | 0 < r < R\}$ with $\sigma^{2p-2} = p(p-1)^{-2}$. On D_R , w_R satisfies $w_R'' \leq r^{-2p+2} w_R^{2p-1}$. Since $u(r) > br$ for large r , there is some R such that $w_R(R/2) < u(R/2)$. Consider $m(r) \equiv u(r) - w_R(r)$ on D_R and let $E = \{r | m(r) > 0, 0 < r < R\}$. Clearly, (i) m is continuous and $m'' > 0$ on E ; (ii) E is non-empty and open; (iii) $\bar{E} \subset D_R$ [since $u(r) = 0$ for $r < R_0$ and $w_R(r) \rightarrow \infty$ as $r \rightarrow R$]. Therefore m has its maximum on ∂E , but $m = 0$ on ∂E . Therefore E is empty and we have a contradiction.

Thus $u \equiv 0$, which implies that $f \equiv 0 \Rightarrow h \equiv 0 \Rightarrow \psi(x) \leq |x|^b$. We have to prove that $g(x) < 0$ everywhere. The Harnack inequality argument used in Theorem 1 implies that either $g \equiv 0$ or $g(x) < 0$, all x . Since $g(x) < 0$ for x near zero, the result is proved. $\square$

The following complements Theorem 3.

Theorem 4. Suppose $\psi(x) > 0$ satisfies (1.10) on $\mathbb{R}^3$ and ψ is spherically symmetric. If $p < 2$ then $\psi(x) \geq (A + |x|)^{-1/(2p-2)}$ with $A = 1/(p-1)$.

Remark. The hypothesis that ψ is spherically symmetric is not really necessary but it simplifies the proof. It is assumed here because Theorem 5 states that any positive solution is necessarily spherically symmetric when $3/2 \leq p < 2$.

Proof. Let $f(r) = (A + r)^b$ with $b = -1/(2p-2)$. Then $\Delta f = Wf$ with $W(r) = b(b-1)(A+r)^{-2} + (2b/r)(A+r)^{-1}$. $W(r) > f(r)^{2p-2} - r^{-1}$. Let $B = \{r > 0 | \psi(r) < f(r)\}$, which is open and which is assumed to be non-empty. Let D be a connected component of B . Thus $D = (s, t)$ with $0 \leq s < t$. Let $h = f - \psi > 0$ on D . On D , $W(r) > \psi(r)^{2p-2} - r^{-1}$, so

$$(\Delta - W)h(r) > 0, \quad \text{all } r \in D. \quad (2.8)$$

First, assume t is finite. Multiply (2.8) by f and integrate over the shell $s < |x| < t$. An integration by parts gives

$$K(t) - K(s) > 0 \quad (2.9)$$

with $K(r) = r^2 f(r)^2 (h/f)'(r)$. $(h/f)'(t) \leq 0$ since $h(t) = 0$ and $h(r) > 0$ for $s < r < t$. Thus $K(t) \leq 0$. Likewise, if $s > 0$ $K(s) \geq 0$ and therefore (2.8) is a contradiction. If $s = 0$ then $K(s) = 0$ since ψ' and f' are bounded near $r = 0$.

Now assume $t = \infty$. Integrate (2.9) over $s < |x| < T$. (2.8) holds as before (with T) and $K(s) \geq 0$. Since $p < 2$, $r f(r)^2 \rightarrow 0$ as $r \rightarrow \infty$. $0 < (h/f)(r) < 1$ on D , so there is a sequence $\{T_n\}$, with $T_n \rightarrow \infty$, such that $T_n (h/f)'(T_n) \rightarrow 0$. Thus $K(T_n) \rightarrow 0$ and (2.8) is again a contradiction. $\square$

Theorem 5. *There is at most one $\psi \neq 0$ that satisfies (1.10) on $\mathbb{R}^3$ in the sense of distributions with the properties that $\psi(x) \geq 0$, all x and $\psi \in L^6(\mathbb{R}^3)$. (By Theorem 3, the second property is automatic if $3/2 \leq p < 2$.)*

Remark. It is not at all clear whether or not $p \geq 2$ is special. If it is additionally assumed that $\nabla \psi \in L^2(\mathbb{R}^3)$ then by Theorem 3 and a Sobolev inequality, $\psi \in L^6$ when $p \geq 3/2$. In Sect. III it will be shown that there is a positive solution to (1.10) for $3/2 < p < 2$, and this solution indeed has $\nabla \psi \in L^2(\mathbb{R}^3)$. However, the methods of Sect. III are not applicable for $p \geq 2$.

Proof. Let ψ satisfy (1.10) with the stated properties. By Theorem 1, ψ is continuous and $\psi(x) > 0$, all x . If $g \in C_0^\infty$, then $-\int \frac{\Delta \psi}{\psi} |g|^2 \leq \int |\nabla g|^2$. (This fact, which was used in [3], follows by integrating by parts and using the Cauchy-Schwarz inequality.) Thus, for any real $g \in C_0^\infty$,

$$L_\psi(g) = \int |\nabla g|^2 + \int W_\psi(x) |g(x)|^2 dx \geq 0, \quad (2.10)$$

where $W_\psi(x) = -|x|^{-1} + \psi(x)^{2p-2}$. Let f_n be a sequence of spherically symmetric C_0^∞ functions with the properties: (i) $0 \leq f_n(x) \leq 1$; (ii) $f_n(x) = 1$ for $2/n < |x| < n$; (iii) $f_n(x) = 0$ for $0 \leq |x| \leq 1/n$ and $|x| \geq 2n$; (iv) $\int |\nabla f_n|^3 \leq A$ for some fixed A . Such a sequence is easy to construct. Let $g_n(x) = f_n(x) \psi(x)$. $g_n \in C_0^\infty$ by Theorem 1. $\int |\nabla g_n|^2 = -\int f_n^2 \psi \Delta \psi + \int \psi^2 |\nabla f_n|^2$. Thus, $L_\psi(g_n) = \int \psi^2 |\nabla f_n|^2 \equiv T_n$. Let $\chi_n(x) = 0$ if $2/n \leq |x| \leq n$ and $\chi_n(x) = 1$ otherwise. Then $T_n = \int \chi_n \psi^2 |\nabla f_n|^2 \leq \|\chi_n \psi\|_6^2 \|\nabla f_n\|_3^2$. Since $\chi_n \rightarrow 0$ pointwise and $\psi \in L^6$, $T_n \rightarrow 0$ as $n \rightarrow \infty$.

Now let ψ_1 and ψ_2 be two different solutions to (1.10) with the stated properties. Denote the corresponding two functionals in (2.10) by L_1 and L_2 and the corresponding sequences by $g_n^1 = f_n \psi_1$ and $g_n^2 = f_n \psi_2$. Then $0 \leq L_1(g_n^2) + L_2(g_n^1) = L_1(g_n^1) + L_2(g_n^2) - \int (\psi_1^{2p-2} - \psi_2^{2p-2})(\psi_1^2 - \psi_2^2) f_n^2$. Since $L_1(g_n^1)$ and $L_2(g_n^2) \rightarrow 0$, the right side of this inequality is negative for large n if $\psi_1 \neq \psi_2$. $\square$

Remark. If $\psi \geq 0$ is unique then ψ is obviously spherically symmetric. Therefore the following theorem is applicable if $3/2 \leq p < 2$.

Theorem 6. *If $\psi \geq 0$ satisfies (1.10), the bound of Theorem 3, and if ψ is spherically symmetric, then $\psi(r)$ is a strictly decreasing function of r .*

Proof. $\Delta \psi \leq 0$ so ψ is superharmonic. Therefore the minimum of $\psi(x)$ in the set $|x| \leq R$ must occur at R . If $\psi(r) = \psi(R)$, $r < R$, then $\psi(x)$ is a constant for $r \leq |x| \leq R$, but this does not satisfy (1.10). $\square$

III. Existence of a Positive Solution of Eq. (1.10) for $3/2 < p < 2$

In Sect. II it was shown that any positive distributional solution to (1.10) with $\psi \in M$ has certain nice properties, especially when $p \geq 3/2$. If $3/2 \leq p < 2$ the solution is unique. In this section it will always be assumed that $3/2 < p < 2$ (note the inequality $p > 3/2$). A positive solution will be shown to exist on $\mathbb{R}^3$ under this condition.

An interesting open question is whether (1.10) has solutions, positive or otherwise, when $p \leq 3/2$ or $p \geq 2$. The method given here sheds no light on this question.

Consider the functional F_p given by (1.8), (1.11) and the class of measurable functions

$$G_p = \{\psi \mid \nabla \psi \in L^2, F_p(\psi) < \infty\}. \quad (3.1)$$

G_p is not empty because

$$f_p(x) = (1 + |x|^2)^{-1/4(p-1)} \quad (3.2)$$

is in G_p for $3/2 < p < 2$. We also define

$$W_p(\psi) = \int k_p(\psi(x), x) dx \geq 0, \quad (3.3)$$

$$E_p = \inf\{F_p(\psi) \mid \psi \in G_p\}. \quad (3.4)$$

Remark. The condition $3/2 < p < 2$ results from the requirement that G_p be non-empty. Note that $k_p(\psi, x)$ has the form $|x|^{-p/(p-1)} h_p(|\psi| |x|^{1/(2p-2)})$ and that $h_p(a) = 0$ if and only if $a = 1$. Suppose $p \geq 2$. If $k_p(\psi(x), x)$ is to be integrable at infinity then $\psi(x) \sim |x|^{-1/(2p-2)}$. But then $|\nabla \psi|^2$ is not integrable at infinity. If $p \leq 3/2$ then, for a similar reason, $\psi(x) \sim |x|^{-1/(2p-2)}$ near $x=0$, but then $|\nabla \psi|^2$ is not integrable.

Theorem 7. *Let $3/2 < p < 2$. There exists $\psi \in G_p$ such that (i) $F_p(\psi) = E_p$; (ii) $\psi(x) \geq 0$; (iii) ψ satisfies (1.10) in the sense of distributions.*

Proof. Let ψ_n be a minimizing sequence for $F_p(\psi)$. $W_p(|\psi|) = W_p(\psi)$ and $\int |\nabla \psi|^2 \geq \int |\nabla |\psi||^2$, so we can assume $\psi_n(x) \geq 0$, all x . Let $b = -(2p-2)^{-1}$. For $\psi \geq 0$, $k_p(\psi, x) \geq C_p |x|^{-1} [\psi - |x|^b]^2$ with $C_p > 0$. Therefore, if $\psi \in G_p$ then $|x|^{-1/2} \psi - |x|^{b-1/2} \in L^2$. But $|x|^{b-1/2} \in L^2_{\text{loc}}$ so $|x|^{-1/2} \psi \in L^2_{\text{loc}} \Rightarrow \psi \in L^2_{\text{loc}} \Rightarrow \psi \in W^{1,2}(B)$ for any bounded ball, B . Since $F_p(\psi_n) \rightarrow E_p$, $\nabla \psi_n$ is bounded in $L^2(\mathbb{R}^3)$ and ψ_n is bounded in $W^{1,2}(B)$. By the Rellich-Kondrachov theorem [1], a bounded set in $W^{1,2}(B)$ is compactly imbedded in $L^2(B)$. By passing to a subsequence we can assume ψ_n has a limit, ψ , in $L^2(B)$. By taking a further subsequence, we can assume $\psi_n \rightarrow \psi$ pointwise. This can be done for every B , so we can assume $\psi_n \rightarrow \psi$ pointwise in $\mathbb{R}^3$. By passing to a further subsequence we can assume, using the Banach-Alaoglu theorem, that $\nabla \psi_n \rightarrow f$ weakly in $L^2(\mathbb{R}^3)$. Clearly, $f = \nabla \psi$. Therefore $\liminf \int |\nabla \psi_n|^2 \geq \int |\nabla \psi|^2$ and, by Fatou's lemma, $\liminf \int k_p(\psi_n(x), x) \geq \int k_p(\psi(x), x)$. Thus $F_p(\psi) \leq E_p$, so ψ minimizes.

Now $k_p(\psi(x), x) \in L^1$. By the above $|x|^{-1} \psi^2 \in L^1_{\text{loc}} \Rightarrow |x|^{-1} \psi \in L^1_{\text{loc}}$. Since $k_p(\psi(x), x) \in L^1$, $\psi^{2p} \in L^1_{\text{loc}}$, and thus $\psi^{2p-1} \in L^1_{\text{loc}}$. Hence, the right side of (1.10) is a distribution.

Let $\eta \in C_0^\infty$ and replace ψ by $\psi_t = \psi + t\eta$. Let $E(t) = E_p(\psi_t)$, whence $dE/dt = 0$ at $t = 0$. $\int |\nabla \psi_t|^2$ is differentiable with derivative $2 \int \nabla \psi \cdot \nabla \eta$ at $t = 0$. By dominated convergence, $W_p(\psi_t)$ can be differentiated under the integral sign, whence $0 = \int \nabla \eta \cdot \nabla \psi + \int \eta \{ |\psi|^{2p-2} - |x|^{-1} \} \psi$. This is Eq. (1.10) in the sense of distributions. $\square$

The ψ given by Theorem 7 is unique (Theorem 5). ψ satisfies two sum rules, one of which arises from the fact that ψ minimizes F_p . Let us define the following integrals:

$$I_1 = \int |\nabla \psi(x)|^2 dx, \quad (3.5)$$

$$I_2 = \int \{ |x|^{-p/(p-1)} - \psi(x)^{2p} \} dx, \quad (3.6)$$

$$I_3 = \int |x|^{-1} \{ |x|^{-1/(p-1)} - \psi(x)^2 \} dx. \quad (3.7)$$

I_1 is finite since $\psi \in G_p$. I_2 and I_3 are finite (and positive) since $|x|^{-1/(2p-2)} > \psi(x) > (A + |x|)^{-1/(2p-2)}$, Theorems 3 and 4, and since $|x|^{-p/(p-1)} \in L_{\text{loc}}^1$. Clearly,

$$F_p(\psi) = I_1 - p^{-1}I_2 + I_3. \quad (3.8)$$

The physical interpretation of these integrals is the following: I_1 is the gradient contribution to the kinetic energy. I_2/p is the decrease in the “fermionic” part of the kinetic energy relative to the TF value. I_3 is the increase in the electron-nucleus potential energy relative to the TF value.

Theorem 8. *Let ψ be the minimizing ψ of Theorem 7. Then*

$$I_1 : I_2 : I_3 = 2(2-p)(p-1) : p(3-p) : p^2 - 3p + 4.$$

In particular, for $p = 5/3$,

$$I_1 : I_2 : I_3 = 1 : 5 : 4.$$

Proof. If (1.10) is multiplied by ψ and integrated, we find $I_2 = I_1 + I_3$. Next, consider $\psi_t(x) \equiv t^{1/(2p-1)} \psi(tx)$ for $t > 0$. Since $g(t) \equiv F_p(\psi_t)$ has its minimum at $t = 1$, $dg(t)/dt = 0$ at $t = 1$. But $I_1(\psi_t) = t^{-1}I_1$, $I_2(\psi_t) = t^{(4-2p)/(p-1)}I_2$ and $I_3(\psi_t) = t^{(4-2p)/(p-1)}I_3$. The rest is algebra. $\square$

Remark. When $p = 5/3$, Theorem 8 implies the *virial theorem*. The change in kinetic energy is $\delta T = I_1 - p^{-1}I_2$. The change in potential energy is $\delta V = I_3$. Therefore $-2\delta T = \delta V$, as usual.

IV. Asymptotic Expansion for Large r ($3/2 < p < 2$)

We shall be concerned here with the unique, positive solution to (1.10) for $3/2 < p < 2$. By Theorems 3 and 4, $\psi(r) = r^b + O(r^{b-1})$ with $r = |x|$ and $b = -1/(2p-2)$. This suggests an asymptotic expansion of the form

$$\psi(r) = r^b \left\{ 1 + \sum_{j=1}^{\infty} a_j r^{-j} \right\}. \quad (4.1)$$

The coefficients a_j are determined as follows. If (4.1) is inserted into the right side of (1.10) and then expanded, the coefficient of r^{b-1} is zero and the coefficient of $r^{b-j-1} (j \geq 1)$ is of the form

$$P_j(A_{j-1}) + (2p-2)a_j, \quad (4.2)$$

where $A_j \equiv (a_1, \dots, a_j)$ and P_j is a polynomial (with $P_1 \equiv 0$). The coefficient of r^{b-j-1} on the left side is zero for $j=0$ and is

$$(b-j+2)(b-j+1)a_{j-1} \quad (4.3)$$

for $j \geq 1$, with $a_0 \equiv 1$. Equating (4.2) and (4.3),

$$a_j = (2p-2)^{-1} [(b-j+2)(b-j+1)a_{j-1} - P_j(A_{j-1})]. \quad (4.4)$$

Thus, a_j is determined recursively by $a_1, \dots, a_{j-1}$. The first three terms of ψ are thus

$$\psi(r) = r^b \{ 1 - r^{-1}(2p-3)(2p-2)^{-3} - \frac{1}{2}r^{-2}(2p-3)(2p-1)^2(2p-2)^{-6} + O(r^{-3}) \}. \quad (4.5)$$

The correctness of (4.1), with the rule (4.4), will be proved here. The chief difficulty is that Δ is a singular perturbation: The term $a_j r^{b-j}$ in ψ generates a term r^{b-j-1} on the right side of (1.10), but it generates r^{b-j-2} on the left. While $\Delta\psi$ thus appears to be relatively small, it is not really small because its coefficients (4.3) grow as j^2 . For this reason the series (4.1) is probably not convergent. In the proof of Theorem 9, $\Delta\psi$ is controlled by combining it with the leading term in (4.2), namely $(2p-2)r^{-1}\psi$.

Theorem 9. *The asymptotic expansion (4.1) and (4.4) is correct, i.e. for any J ,*

$$\psi(r) = r^b \left\{ 1 + \sum_{j=1}^J a_j r^{-j} \right\} + o(r^{b-J}) \quad (4.6)$$

as $r \rightarrow \infty$.

Proof. The proof is by induction. Theorems 3 and 4 assert that (4.6) is true for $J=0$. Assuming (4.6) holds for some $J \geq 0$, we will prove that (4.6) holds for $J+1$.

Write $\psi = \phi + g$ with $\phi(r) = r^b \left\{ 1 + \sum_{j=1}^J a_j r^{-j} \right\}$. For any $\varepsilon > 0$ there is an R_ε such that for all $r > R_\varepsilon$: (i) $\phi(r) > 0$, (ii) $g(r)/\phi(r) < \varepsilon$, (iii) $[\phi(r) + g(r)]^{2p-1} - \phi(r)^{2p-1} \equiv U(r)g(r)$ with $|U(r) - (2p-1)r^{-1}| < \varepsilon r^{-1}$.

Equation (1.10) reads

$$[\Delta + W(r)]g(r) = h(r), \quad (4.7)$$

with $h = \phi^{2p-1} - r^{-1}\phi - \Delta\phi$ and $W(r) = r^{-1} - U(r)$. Let us examine (4.7) for $r > R_\varepsilon$. As $r \rightarrow \infty$, $h(r) = Kr^{b-J-2} + o(r^{b-J-2})$ since $a_1, \dots, a_J$ satisfy (4.4). Moreover, $K = -(2p-2) \cdot [\text{right side of (4.4) with } j=J+1]$. We also know that $g(r) = o(r^{b-J})$ by the induction hypothesis. Finally, $|W(r) + (2p-2)r^{-1}| < \varepsilon r^{-1}$ by (iii) above. The point of writing (1.10) as (4.7) is that we now regard W as a fixed function that is close to $-(2p-2)r^{-1}$. It is true that W “depends on g ”, but that information is suppressed.

Lemma 10 will imply that $g(r) = -K(2p-2)^{-1}r^{b-J-1} + o(r^{b-J-1})$, which is the desired result. To see this, let $z > 0$ and define $v_z(r) = \exp[-2(zr)^{1/2}]$. Let $W_z(r) = -z/r$. Then $(\Delta + W_z)v_z = -\frac{3}{2}z^{1/2}r^{-3/2}v_z$. Without loss, assume $K \geq 0$ [otherwise replace g by $-g$ in (4.7)]. Let $z_0 = 2p-2$.

Lower Bound for g . Pick $0 < \varepsilon < z_0$ and let $z = z_0 - \varepsilon$. For $r > R_\varepsilon$, $W(r) \leq -z/r = W_z(r)$. Pick $\delta > 0$ and $A \geq 0$ and let $f(r) = -Av_z(r) - (K + \delta)z^{-1}r^{b-J-1} < 0$. Then

$$(\Delta + W_z)f(r) = \frac{3}{2}Az^{1/2}r^{-3/2}v_z(r) + (K + \delta)r^{b-J-2} - (K + \delta)z^{-1}(b-J)(b-J-1)r^{b-J-3}.$$

For any fixed $\varepsilon, \delta > 0$ we can choose $A \geq 0$ such that (i) $(\Delta + W_z)f \geq h = (\Delta + W)g$ for $r > R_\varepsilon$, (ii) $g(R_\varepsilon) \geq f(R_\varepsilon)$. By Lemma 10 [with $B = \{r | r > R_\varepsilon\}$, $g_1 = g$, $g_2 = f$, $W_1 = W$, $W_2 = W_z$, and the facts that $g(r)$ and $f(r) \rightarrow 0$ as $r \rightarrow \infty$ and $f(r) < 0$], $g(r) \geq f(r)$ for $r > R_\varepsilon$. This implies $\liminf_{r \rightarrow \infty} r^{-b+J+1}g(r) \geq -(K + \delta)/z$. Since δ and ε are arbitrary, $\liminf_{r \rightarrow \infty} r^{-b+J+1}g(r) \geq -K/z_0$.

Upper Bound for g . This is similar to the preceding. For $r > R_\varepsilon$, $-z/r \geq W(r) \geq -q/r$ with $q = z_0 + \varepsilon$ and $z = z_0 - \varepsilon$. Let $G(r) = g(r) - Av_z(r) - \delta r^{b-J-1}$. Then

$$(\Delta + W)G(r) = h(r) + \frac{3}{2}Az^{1/2}r^{-3/2}v_z(r) - A(W(r) + zr^{-1})v_z(r) - \delta(b-J)(b-J-1)r^{b-J-3} - \delta W(r)r^{b-J-1}.$$

Let $F(r) = -Kq^{-1}r^{b-J-1} \leq 0$, whence

$$(\Delta + W_q)F(r) < Kr^{b-J-2}.$$

For any fixed $\varepsilon, \delta > 0$ we can choose $A \geq 0$ such that (i) $(\Delta + W)G \geq (\Delta + W_q)F$ for $r > R_\varepsilon$, (ii) $F(R_\varepsilon) \geq G(R_\varepsilon)$. By Lemma 10 ($g_1 = F$, $g_2 = G$, $W_1 = W_q$, $W_2 = W$), $F(r) \geq G(r)$. As before this implies $\limsup_{r \rightarrow \infty} r^{-b+J+1}g(r) \leq -K/z_0$.

These two bounds yield the desired result. $\square$

Lemma 10. Let $B \subset \mathbb{R}^3$ be an open set, let g_i , $i=1,2$, be two functions on $\mathbb{R}^3$ which are continuous on some neighborhood of $\bar{B}$, the closure of B , and let W_i be two functions in $L^1_{\text{loc}}(B)$. Suppose that $[\Delta + W_1]g_1 \leq [\Delta + W_2]g_2$ as distributions on B . In addition suppose that (i) $g_1 \geq g_2$ on the boundary of B , (ii) if B is unbounded, then for every $\varepsilon > 0$ there exists an R_ε such that $g_1(x) - g_2(x) \geq -\varepsilon$ on $\{x | x \in B, |x| \geq R_\varepsilon\}$, (iii) $0 \geq W_2(x) \geq W_1(x)$, all $x \in B$, (iv) $g_1(x) \leq 0$ whenever $g_1(x) < g_2(x)$. Then $g_1(x) \geq g_2(x)$ for all $x \in B$.

Proof. Let $\phi = g_1 - g_2$. Let $D = \{x | x \in B, \phi(x) < 0\}$, which is open. Then $\Delta\phi \leq W_2g_2 - W_1g_1$, and thus $\Delta\phi \leq 0$ on D by (iii) and (iv). Let $U_\varepsilon = \{x | |x| < R_\varepsilon\}$. ϕ is superharmonic on $D \cap U_\varepsilon$, but $\phi = 0$ on ∂D and $\phi \geq -\varepsilon$ on ∂U_ε . Thus $\phi \geq -\varepsilon$ on $D \cap U_\varepsilon$. Since ε is arbitrary and R_ε can be chosen to tend to infinity as $\varepsilon \rightarrow 0$, $\phi \geq 0$ on D , so D is empty. $\square$

Acknowledgement. It is a pleasure to thank Prof. Kenji Yajima for a helpful discussion.

References

1. Adams, R.A.: Sobolev spaces. New York: Academic Press 1975
2. Benguria, R.: The von Weizsäcker and exchange corrections in Thomas-Fermi theory. Ph. D. thesis, Princeton University 1979 (unpublished)
3. Benguria, R., Brezis, H., Lieb, E.H.: The Thomas-Fermi-von Weizsäcker theory of atoms and molecules. *Commun. Math. Phys.* **79**, 167–180 (1981)
4. Fermi, E.: Un metodo statistico per la determinazione di alcune priorieta dell'atome. *Rend. Accad. Naz. Lincei* **6**, 602–607 (1927)
5. Gilbarg, D., Trudinger, N.: Elliptic partial differential equations of second order. Berlin, Heidelberg, New York: Springer 1977
6. Kato, T.: On the eigenfunctions of many particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **10**, 151–171 (1957)
7. Kato, T.: Schrödinger operators with singular potentials. *Isr. J. Math.* **13**, 135–148 (1973)
8. Liberman, D., Lieb, E.H.: Numerical calculation of the Thomas-Fermi-von Weizsäcker function for an infinite atom without electron repulsion, Los Alamos National Laboratory report (in preparation)
9. Lieb, E.H.: Thomas-Fermi and related theories of atoms and molecules. *Rev. Mod. Phys.* **53**, 603–641 (1981)
10. Lieb, E.H., Simon, B.: The Thomas-Fermi theory of atoms, molecules, and solids. *Adv. Math.* **23**, 22–116 (1977)
11. Morrey, C.B., Jr.: Multiple integrals in the calculus of variations. Berlin, Heidelberg, New York: Springer 1966
12. Stampacchia, G.: Equations elliptiques du second ordre à coefficients discontinus. Montréal: Presses de l'Univ. 1965
13. Thomas, L.H.: The calculation of atomic fields. *Proc. Camb. Phil. Soc.* **23**, 542–548 (1927)
14. von Weizsäcker, C.F.: Zur Theorie der Kernmassen. *Z. Phys.* **96**, 431–458 (1935)

Communicated by R. Jost

Received November 11, 1981

The most negative ion in the Thomas–Fermi–von Weizsäcker theory of atoms and molecules

Rafael Benguria^{†§} and Elliott H Lieb^{‡||}

[†] Departamento de Física, Universidad de Chile, Casilla 5487, Santiago, Chile

[‡] Departments of Mathematics and Physics, Princeton University, PO Box 708, Princeton, NJ 08544, USA

Received 4 May 1984

Abstract. Let N_c^- denote the maximum number of electrons that can be bound to an atom of nuclear charge z , in the Thomas–Fermi–von Weizsäcker theory. It is proved that N_c^- cannot exceed z by more than one, and thus this theory is in agreement with experimental facts about real atoms. A similar result is proved for molecules, i.e. N_c^- cannot exceed the total nuclear charge by more than the number of atoms in the molecule.

1. Introduction

The Thomas–Fermi–von Weizsäcker (TFW) theory (von Weizsäcker 1935, Benguria *et al* 1981, Lieb 1981) is defined by the energy functional (see Lieb 1981, § VII) (in units in which $\hbar^2(2m)^{-1} = |e| = 1$, where e and m are the electron charge and mass)

$$\xi(\rho) = A \int (\nabla \rho^{1/2}(x))^2 dx + \frac{3}{5} \gamma \int \rho(x)^{5/3} dx - \int V(x) \rho(x) dx + D(\rho, \rho) \quad (1)$$

where

$$D(\rho, \rho) = \frac{1}{2} \int \rho(x) |x - y|^{-1} \rho(y) dx dy \quad (2)$$

$$V(x) = \sum_{j=1}^K z_j |x - R_j|^{-1}. \quad (3)$$

Here $z_1, z_2, \dots, z_K \geq 0$ are the charges of K fixed nuclei located at $R_1, \dots, R_K$. The total nuclear charge is denoted by Z , $Z = \sum_{j=1}^K z_j$. $K = 1$ is the atomic case and here we shall simply write $Z = z_1 = z$. dx is always a three-dimensional integral. $\xi(\rho)$ is defined for electronic densities $\rho(x) \geq 0$ such that each of the terms of $\xi(\rho)$ in (1) is finite. In the physical situation, $\gamma = \gamma_{\text{phys}} = (3\pi^2)^{2/3}$ but, for generality, we shall allow γ to be an arbitrary positive constant in what follows. The TFW energy for N (not necessarily an integer) electrons is defined by

$$E(N) = \inf \left(\xi(\rho) \mid \int \rho = N \right). \quad (4)$$

[§] Work partially supported by Dpto Desarrollo Investigación, Universidad de Chile.

^{||} Work partially supported by US National Science Foundation grant PHY-8116101-A02.

On energetic grounds, the value of A should be chosen to reproduce the Scott term in the expression of the atomic or molecular energy $E(N)$ as a function of N and the nuclear charges, (see Lieb 1981, §§ V.B, VII.D). Numerically one finds (Lieb and Liebman 1982, Lieb 1982), $A = 0.1859$. However, we should retain A as an arbitrary positive constant. In the original TFW model (von Weizsäcker 1935) the numerical value of A is 1.

It is known (Benguria *et al* 1981, Lieb 1981) that there exists a critical value of N (depending on A , γ and the z_j and R_j), which we denote by N_c , such that for $N \leq N_c$ the minimisation problem (4) has a unique solution, whereas for $N > N_c$ there is no solution. In other words, N_c is the maximum number of ‘electrons’ that can be bound to the atom or molecule. The aim of this paper is to find an upper bound for N_c . The value of N_c is given by $\int \rho$, where $\rho \geq 0$ is the unique minimising function of $\xi(\rho)$ without constraints. Let $\psi = \rho^{1/2}$. Then ψ is the *unique* positive solution of the TFW equation (for a saturated system),

$$-A \Delta \psi(x) + (\gamma \psi(x)^{4/3} - \phi(x)) \psi(x) = 0 \quad (5)$$

where

$$\phi(x) = V(x) - |x|^{-1} * \rho \quad \text{with } \rho = \psi^2. \quad (6)$$

Note that (5) is the Euler equation corresponding to the functional $\xi(\psi^2)$. The only previous rigorous results (Lieb 1981, Benguria *et al* 1981) for N_c were that $Z < N_c < 2Z$.

Our main result is the following.

Theorem 1. For a TFW molecule of $K \geq 1$ atoms,

$$0 < N_c - Z \leq 270.74(A/\gamma)^{3/2} K \quad (7)$$

for all choices of $z_1, \dots, z_K$ and $R_1, \dots, R_K$. In particular, for the value of A chosen in Lieb (1981) and Lieb and Liebman (1982) to reproduce the Scott term in the energy (i.e. $A = 0.1859$) and for the physical $\gamma = (3\pi^2)^{2/3}$,

$$0 < N_c - Z < 0.7335K. \quad (8)$$

In the TFW model the number of electrons is not generally an integer, but in a real atom N and z are required to be integral. How can theorem 1 be interpreted in the light of this additional requirement? One way is the energetic point of view: since $E(N)$ is strictly decreasing for $N < N_c$ and constant for $N \geq N_c$ (Lieb 1981, § VII.A), theorem 1 implies that $E(z) > E(z+1) = E(z+2)$. Thus, the $(z+1)$ th electron has a positive binding energy, while the $(z+2)$ th does not, and we can say that a singly ionised atom (but not a doubly ionised atom) is stable. This interpretation, however, suffers from the drawback that there is no solution to (5) when $N = z+1$. A second interpretation that leads to the same conclusion about atomic ionisation, but eliminates the problem that (5) has no solution for $N = z+1$, was kindly provided by John Morgan: introduce the Fermi–Amaldi correction (i.e. replace $D(\rho, \rho)$ in (1) by $(1 - 1/N)D(\rho, \rho)$). This has the effect of replacing z by $z'(N) = Nz/(N-1)$. (It also effectively changes A and γ , but not A/γ .) Theorem 1 now states that a solution to (5) always exists if $N \leq z'(N)$ while it never exists if $N - 0.74 \geq z'(N)$. This implies that a solution exists (with N and z integral) if and only if $N \leq z+1$. However it is not clear that $E(z+1) < E(z)$ in this Fermi–Amaldi model.

The best previous upper bound on N_c is, as we said, $N_c < 2Z$ (Lieb 1981, theorem 7.23), a result which is valid for both atoms and molecules. It turns out that such a

bound also holds for the Hartree (bosonic) atom. More recently one of us (Lieb 1984a, Lieb *et al* 1984) has proved a similar bound for the real Schrödinger equation namely, $N_c < 2z + 1$ for an atom and $N_c < 2Z + K$ for a molecule of K atoms. This result (Lieb 1984a, Lieb *et al* 1984) is valid regardless of the statistics of the bound particles. However, if the bound particles are fermions, as is the case for real matter, N_c should presumably not exceed z (for an atom) by more than one or two electrons. This is still a conjecture; however, it has been proved that $N_c/z \rightarrow 1$ as $z \rightarrow \infty$ for fermions (Lieb *et al* 1984). On the other hand, we know that $N_c - z > 0.2z$ for the Schrödinger equation of an atom with bosonic particles and for large z (Benguria and Lieb 1983, Baumgartner 1983, 1984). See (Lieb 1984b) for a review of the recent literature on the subject.

In the Thomas–Fermi theory, defined by the energy functional (1) with $A=0$, N_c is exactly Z even in the molecular case (Lieb 1981, theorem 3.18, Lieb and Simon 1977). Equation (7) implies that for the TFW atom or molecule $N_c \rightarrow Z$ as $A \rightarrow 0$. However, we do not expect $N_c(A)$ to be analytic around $A=0$ because the von Weizsäcker correction is a singular perturbation to Thomas–Fermi theory. It is an *open problem* to derive an asymptotic expansion for $N_c(A)$ around $A=0$.

Two other open problems arise from the results of this paper. The first is that while we prove an upper bound for $N_c - Z$, we have no *lower* bound. We conjecture that $N_c - Z \rightarrow \text{constant} > 0$ as $Z \rightarrow \infty$. The second problem is related to the first: it is highly plausible that $N_c - Z$ is a monotonically increasing function of all the z_j (for fixed $R_1, \dots, R_K$). Is this true?

This article is organised as follows: in § 2 we give the proof of theorem 1; in § 3 we determine the behaviour of N_c as Z goes to zero. Finally in § 4 we give a bound for the chemical potential of a neutral molecule. Such a bound is independent of the charge of the nuclei.

We should like to emphasise that many of the results herein can be extended in two ways: (i) to spherically symmetric ‘smeared out’ nuclei; (ii) to the TFW theory in which the exponent $\frac{5}{3}$ in (1) is replaced by some $p \neq \frac{5}{3}$ (cf Lieb 1981). For simplicity and clarity we confine ourselves here to point nuclei and $p = \frac{5}{3}$.

2. Proof of theorem 1

The proof of theorem 1 will be divided into three steps. First, we estimate the excess charge $Q \equiv N_c - Z$ in terms of the electronic density ρ and the TFW potential ϕ evaluated at an arbitrary, but fixed, distance r from all the nuclei. Then we find a local bound for ρ in terms of ϕ . These two estimates do not involve the z_j explicitly. Therefore, if we can prove that at some distance of order one, (i.e. independent of the z_j) the potential ϕ is bounded by a constant independent of the z_j , then the two previous results will imply that Q is less than a constant independent of the z_j , which is basically what theorem 1 says. Proving this last fact about ϕ is our third step. We begin with

Lemma 2. Let $\psi \geq 0$, ϕ be the unique solution pair for the TFW equation (5), (6) with V being the potential (3) for a molecule. Then, the function

$$p(x) = (4\pi A \psi(x)^2 + \phi(x)^2)^{1/2} \quad (9)$$

is subharmonic away from the nuclei, i.e. on $\mathcal{R}^3 \setminus \bigcup_{j=1}^K R_j$.

1048 *R Benguria and E H Lieb*

Proof. By direct computation

$$\Delta p = p^{-1}(4\pi A\psi \Delta\psi + \phi \Delta\phi) + h \quad (10)$$

with

$$h = 4\pi A p^{-3} |\phi \nabla \psi - \psi \nabla \phi|^2 \geq 0.$$

By (5), (6) the sum of the first two terms in (10) is (away from the nuclei so that $\Delta V = 0$): $4\pi\gamma\psi^{10/3} \geq 0$. Thus, $\Delta p \geq 0$, so p is subharmonic. $\square$

Remark. Let

$$W(x) \equiv \gamma\psi(x)^{4/3} - \phi(x) \quad (11)$$

be the ‘potential’ in the TFW equation (5). Proceeding as in the proof of lemma 2, one can show that

$$(4\pi A\psi(x)^2 + W(x)^2)^{1/2} \quad (12)$$

is subharmonic whenever $W(x) \geq 0$.

The next lemma gives a local bound for ψ in terms of ϕ . This bound is independent of the nuclear charges z_j .

Lemma 3. For all $\lambda \in (0, 1)$ and all $x \in \mathcal{R}^3$

$$\gamma\lambda\psi(x)^{4/3} \leq \phi(x) + c(\lambda)A^2\gamma^{-3} \quad (13)$$

with

$$c(\lambda) \equiv (9/4)\pi^2\lambda^{-2}(1-\lambda)^{-1}. \quad (14)$$

Proof. Define $u(x) \equiv \psi(x)^{4/3}$. Then, from the TFW equation (5),

$$-\Delta u + (4/3A)(\gamma u - \phi)u + |\nabla u|^2/4u = 0$$

and hence

$$-\Delta u + (4/3A)(\gamma u - \phi)u \leq 0. \quad (15)$$

Also, from (6)

$$-\Delta\phi = -4\pi\psi^2 = -4\pi u^{3/2} \quad x \neq R_j, \text{ all } j. \quad (16)$$

Let $v(x) = \gamma\lambda u(x) - \phi(x) - d$, with d a positive constant. We shall show that $v(x) \leq 0$, all x , for appropriate d and λ . From equations (15) and (16),

$$-\Delta v \leq -(4\gamma\lambda/3A)(\gamma u - \phi)u + 4\pi u^{3/2}.$$

Let $S = \{x | v(x) > 0\}$. ψ is continuous and goes to zero at infinity; ϕ is continuous away from the R_j and it also goes to zero at infinity (Benguria *et al* 1981, § III). Therefore v is continuous away from all the R_j and goes to $-d$ at infinity. Hence S is open and bounded. Moreover, $R_j \notin S$, all j since $\phi = +\infty$ at the R_j . On S , $\phi < \gamma\lambda u - d$ so

$$\begin{aligned} -\Delta v &\leq -(4\gamma\lambda/3A)(\gamma u + d - \gamma\lambda u)u + 4\pi u^{3/2} \\ &\leq u[4\pi u^{1/2} - (4/3A)\gamma^2\lambda(1-\lambda)u - (4/3A)\gamma\lambda d] \\ &\leq 0 \end{aligned}$$

provided we choose $\lambda \in (0, 1)$ and $d = \frac{9}{4}(\pi A/\lambda)^2 \gamma^{-3}(1-\lambda)^{-1}$ in order that the quantity in brackets [] be non-negative for all possible (unknown) values of $u(x)$. With that choice of λ and d , v is subharmonic on S . On ∂S $v = 0$, and therefore S is empty. $\square$

Corollary. For all $x \in \mathcal{R}^3$

$$\phi(x) \geq -3^5 2^{-4} \pi^2 A^2 \gamma^{-3} \geq -150 A^2 \gamma^{-3}. \quad (17)$$

If x is such that $\phi(x) \leq 0$, then $p(x) = (4\pi A \psi(x)^2 + \phi(x)^2)^{1/2}$ satisfies

$$p(x) \leq \left(\frac{4}{3}\right)^{3/4} 16 \pi^2 A^2 \gamma^{-3} \leq 196 A^2 \gamma^{-3}. \quad (18)$$

Proof. By (13) and the fact that $\psi(x)^{4/3} \geq 0$, $\phi(x) \geq -c(\lambda) A^2 \gamma^{-3}$ for all $\lambda \in (0, 1)$. Minimising $c(\lambda)$ (at $\lambda = \frac{2}{3}$) gives (17). To prove (18), take $\lambda = \frac{3}{4}$ (which minimises $c(\lambda)/\lambda$), let $a = -\frac{3}{4} c A^2 \gamma^{-3}$ and observe from (13) that

$$p(x)^2 \leq \max_{a \leq \phi \leq 0} [4\pi A (3\gamma/4)^{-3/2} (\phi - a)^{3/2} + \phi^2]. \quad (19)$$

The right-hand side of (19) is convex in ϕ , so its maximum occurs either at $\phi = 0$ or $\phi = a$. $\phi = 0$ prevails and gives (18). $\square$

In our next lemma, starting from ρ and ϕ , we introduce a smeared density $\tilde{\rho}$ and potential $\tilde{\phi}$. We find that $\tilde{\rho}$ and $\tilde{\phi}$ satisfy an inequality resembling the Thomas–Fermi equation for smeared nuclei. Then we use a comparison theorem to get an upper bound for the smeared potential $\tilde{\phi}$ in terms of a universal function (independent of the z_j). Finally, noting that ϕ is subharmonic away from the nuclei, we see that essentially the same bound applies to ϕ . In particular, this lemma says that at distances of order one from all the nuclei, in atomic units, ϕ is of order one and, in any case, independent of the z_j . We note, however, that this bound is not satisfactory both very close and very far from the nuclei. Near the nuclei it diverges too fast. On the other hand, the bound is always positive, whereas ϕ is negative at large distances because $Q = N_e - Z$ is strictly positive.

Lemma 4. Let $\psi \geq 0$, ϕ be the solution of the TFW equation (5), (6) with V given by (3). Choose any $R > 0$ and define

$$\delta \equiv 25 \pi^{-2} \gamma^3 = 2.53 \gamma^3 \quad (20)$$

which is independent of the z_j . Suppose $x \in \mathcal{R}^3$ is such that $|x - R_j| > R$ for all $j = 1, 2, \dots, K$. Then

$$\phi(x) \leq A \pi^2 R^{-2} + \delta \sum_{j=1}^K (|x - R_j| - R)^{-4}. \quad (21)$$

Proof. Let $W = \gamma \rho^{2/3} - \phi$, $\rho \equiv \psi^2$ and consider the Hamiltonian $H = -\Delta + W$. H is a non-negative operator, since its ground state, the TFW function ψ , has zero energy (chemical potential). Therefore for any function $b \in L^2$ with $\nabla b \in L^2$ we have

$$A \int |\nabla b(x)|^2 dx + \int W(x) b(x)^2 dx \geq 0. \quad (22)$$

We shall choose $b(x)$ to be a translate of the normalised ground state, $e(x)$, of the Laplacian on a ball of radius R with Dirichlet boundary condition. That is, let

$e(x) = (2\pi R)^{-1/2} \sin(\pi|x|R^{-1})/|x|$, for $|x| \leq R$ and $e(x) = 0$ otherwise. Clearly, $e(x)$ is spherically symmetric, decreasing and it has compact support. Let $b_x(y) = e(y-x)$ denote the translate of e and define $g(x) = e(x)^2$ and $g_x(y) = g(y-x)$. Let $B = A \int |\nabla b_x(y)|^2 dy$. Clearly B does not depend on x . With this choice of b , $B = (\pi/R)^2 A$. From equation (22) we have,

$$\int W(y)g_x(y) dy \geq -B \quad \text{all } x. \quad (23)$$

Note that $\int W(y)g_x(y) dy = (g * W)(x)$, where an asterisk denotes convolution. Define

$$\tilde{\phi} \equiv \phi * g - B. \quad (24)$$

Since $\phi \in L^{3+\varepsilon} + L^{3-\varepsilon}$, $\varepsilon > 0$, (Benguria *et al* 1981, proof of lemma 7) and $g \in L^p$, all $p \geq 1$, $\tilde{\phi}$ is continuous and goes to $-B$ at infinity (Lieb 1981, lemma 3.1). Using Hölder's inequality, we have for all x

$$(g * \rho^{2/3})(x) \leq [(g * \rho)(x)]^{2/3} \left(\int g(y) dy \right)^{1/3} = [(g * \rho)(x)]^{2/3} \quad (25)$$

where we have used $\int g(y) dy = 1$. Let us also define

$$\tilde{\rho} \equiv g * \rho. \quad (26)$$

From equations (23)–(26) we obtain for all x

$$B \geq (\phi * g)(x) - \gamma(\rho^{2/3} * g)(x) \geq \tilde{\phi}(x) + B - \gamma\tilde{\rho}(x)^{2/3}.$$

In other words

$$\tilde{\phi} \leq \gamma\tilde{\rho}^{2/3}. \quad (27)$$

Notice that ϕ is subharmonic away from the nuclei and that $\tilde{\phi} = g * \phi - B$ with g being spherically symmetric, positive, of total mass one and having support in a ball of radius R . From this it follows easily that

$$\phi(x) \leq \tilde{\phi}(x) + B \quad (28)$$

for all x such that $|x - R_j| > R$ (for all j). Thus, to prove (21) we need a bound on $\tilde{\phi}$.

From equations (6) and (24), using the bound (27) and the fact that the Laplacian commutes with convolution, we compute

$$-(4\pi)^{-1} \Delta \tilde{\phi}(x) = \tilde{V}(x) - \tilde{\rho}(x) \leq \tilde{V}(x) - \gamma^{-3/2} [\tilde{\phi}_+(x)]^{3/2} \quad (29)$$

with

$$\tilde{V}(x) = \sum_{j=1}^K z_j g(x - R_j) \quad (30)$$

and with $\tilde{\phi}_+(x) = \max(\tilde{\phi}(x), 0)$.

Note that equation (29) resembles a Thomas–Fermi (TF) equation with smeared nuclei of spherical charge density $z_j g(x - R_j)$. Indeed, let $\hat{\phi}$ be the TF potential for this system (i.e. with equality in (29)):

$$-(4\pi)^{-1} \Delta \hat{\phi}(x) = \tilde{V}(x) - \gamma^{-3/2} \hat{\phi}(x)^{3/2}. \quad (31)$$

It is known from general TF theory that (31) has a unique solution, $\hat{\phi}$, that goes to zero at infinity.

It is easy to see that

$$\tilde{\phi}(x) \leq \hat{\phi}(x) \quad \text{for all } x \quad (32)$$

by observing that if the set $M = \{x | \hat{\phi}(x) - \tilde{\phi}(x) < 0\}$, then $\hat{\phi} - \tilde{\phi}$ is superharmonic on M and zero on the boundary of M and infinity, so M is empty.

The next step is to bound $\hat{\phi}$. First consider an atom with $V = z/r$, $r = |x|$, and consider the function $f(r) = \delta(r - R)^{-4}$ which satisfies

$$(4\pi)^{-1} \Delta f \leq \gamma^{-3/2} f^{3/2} \quad \text{for } r > R. \quad (33)$$

Outside the ball of radius R (centred at the origin) $\hat{\phi}$ satisfies

$$(4\pi)^{-1} \Delta \hat{\phi} = \gamma^{-3/2} \hat{\phi}^{3/2}. \quad (34)$$

Again, by a comparison argument (and using the fact that $f(r) - \hat{\phi}(r) = \infty$ when $r = R$)

$$\hat{\phi}(r) \leq f(r) \quad \text{for } r > R. \quad (35)$$

This, together with (28), proves (21) in the atomic case.

For the molecular case, let $\hat{\phi}_j(x)$ be the solution to (31) for an atom of (smeared) nuclear charge z_j located at R_j . By another comparison argument (Lieb and Simon 1977, theorem V.12 or Lieb 1981, corollary 3.6), $\hat{\phi}(x) \leq \sum_{j=1}^K \hat{\phi}_j(x)$. This, together with (35) and (28) proves (21). $\square$

We conclude this section with

Proof of Theorem 1: atomic case. Let us start with the atomic case, $V(x) = z/|x|$, in order to expose the ideas most simply. The following facts have been established:

$$p(x) = (4\pi A \psi(x)^2 + \phi(x)^2)^{1/2} \quad (36)$$

is subharmonic for $|x| > 0$.

$$p(x) \leq (4/3)^{3/4} 16\pi^2 A^2 \gamma^{-3} \quad (37)$$

if $\phi(x) \leq 0$.

$$\phi(x) \leq \delta(|x| - R)^{-4} + \pi^2 A R^{-2} \quad (38)$$

for all $|x| > R > 0$, with $\delta = 25\gamma^3\pi^{-2}$ and with arbitrary $R > 0$.

$$\gamma\lambda\psi(x)^{4/3} \leq \phi(x) + c(\lambda)A^2\gamma^{-3} \quad (39)$$

for all $|x|$, $0 < \lambda < 1$ with $c(\lambda) = 9\pi^2[4\lambda^2(1-\lambda)]^{-1}$.

The functions p , ϕ , ψ are functions only of $|x| = r$. As $r \rightarrow \infty$, $\psi(r) \rightarrow 0$ faster than any power of r (Lieb 1981, theorem 7.24) and $r\phi(r) \rightarrow -Q$. Thus,

$$rp(r) \rightarrow Q \quad \text{as } r \rightarrow \infty. \quad (40)$$

The subharmonicity and the fact that $p(r) \rightarrow 0$ as $r \rightarrow \infty$ imply that $rp(r)$ is monotonically decreasing and convex. (This may be seen from the fact that $\Delta p \geq 0$ is equivalent, in polar coordinates, to $d^2(rp(r))/dr^2 \geq 0$.) Using (40) we conclude that

$$Q \leq rp(r) \quad \text{for any } r > 0. \quad (41)$$

The same conclusion, (41), can be reached from another viewpoint, which will be important for the molecular case: fix $r > 0$ and consider the domain $D_r = \{x | |x| > r\}$. Let P be any harmonic function on D_r with $P(x) \rightarrow 0$ as $|x| \rightarrow \infty$ and $P(x) \geq p(x)$ on

the boundary $|x| = r$. Then $P(x) \geq p(x)$ for all $x \in D_r$. (Proof: on the set $E = \{x | P(x) < p(x)\} \subset D_r$, $p - P$ is positive, subharmonic and $p - P$ vanishes on the boundary of E , so E is empty.) Now choose $P(x) = rp(r)/|x|$ for $|x| > r$. Then $rp(r) \geq |x|p(x)$, for $|x| > r$. However, $|x|p(x) \rightarrow Q$ as $|x| \rightarrow \infty$, and this establishes (41).

To complete the proof, we make the following specific choices for r, R, λ :

$$r = 0.9086 (\gamma^3/A)^{1/2} \quad R = 0.4020 (\gamma^3/A)^{1/2} \quad \lambda = 0.7825. \quad (42)$$

If $\phi(r)$ happens to be positive, we use the bound (38), followed by (39) and insert these in (36). (41) then implies that if $\phi(r) > 0$ then

$$Q \leq 270.74 (A/\gamma)^{3/2}. \quad (43)$$

(The numbers in (42) were chosen to minimise the coefficient in (43).) On the other hand, if $\phi(r) \leq 0$ we can use (37) and (41) to conclude that

$$Q \leq 178.03 (A/\gamma)^{3/2}. \quad (44)$$

Clearly, (43) is the worst case, and this gives theorem 1. Note, however, that if it were to be shown that $\phi(r) \leq 0$, then the bound (44) would be valid, and, using the physical values of A and γ , one would obtain $Q \leq 0.49$.

Molecular case. Equation (36) is still valid, except that p is subharmonic only on the set $x \neq R_j$ (all $j = 1, \dots, K$). Equations (37) and (39) are also valid. Equation (38) must be replaced by (21) on the set

$$D_R = \{x | |x - R_j| > R \text{ for all } j = 1, \dots, K\}. \quad (45)$$

Now choose r, R and λ as in (42) and consider the smaller domain

$$D_r = \{x | |x - R_j| > r \text{ for all } j = 1, \dots, K\}. \quad (46)$$

Consider the following function which is harmonic on D_r :

$$P(x) = Q_1 \sum_{j=1}^K |x - R_j|^{-1} \quad (47)$$

where Q_1 is the right-hand side of (43), namely the value of the upper bound for $rp(r)$ computed in the atomic case under the assumption $\phi(r) \geq 0$. As explained above, if we can show that $P(x) \geq p(x)$ for all x on the boundary of D_r , then $P(x) \geq p(x)$ for all $x \in D_r$. Taking the limit $|x| \rightarrow \infty$ yields

$$Q = \lim_{|x| \rightarrow \infty} |x|p(x) \leq \lim_{|x| \rightarrow \infty} |x|P(x) = KQ_1 \quad (48)$$

which is the desired result.

Let x be on the boundary of D_r , so that $|x - R_j| = r$ for some j (say $j = m$). If $\phi(x) \leq 0$, the bound (37) is valid and $p(x) \leq Q_2/r$, where $Q_2 < Q_1$ is the right-hand side of (44). However, $P(x) \geq Q_1|x - R_m|^{-1} = Q_1/r$, so $P(x) > p(x)$. On the other hand, suppose $\phi(x) \geq 0$, in which case we can use (21) and (39). Now use proposition 5 below with the choices $t = \frac{2}{3}$, $s = 2$ and

$$\begin{aligned} a_j &= \delta(|x - R_j| - R)^{-4} & b_j &= a_j(4\pi A)^{2/3}/\gamma\lambda & \text{for } j \neq m \\ a_m &= \delta(|x - R_m| - R)^{-4} + A\pi^2 R^{-2} \\ b_m &= (a_m + c(\lambda)A^2\gamma^{-3})(4\pi A)^{2/3}/\gamma\lambda. \end{aligned}$$

Recalling that $|x - R_m| = r$ we have

$$p(x) \leq p_1(r) + \sum_{j \neq m} \hat{p}(x - R_j) \quad (49)$$

where $p_1(r)$ is precisely the number we calculated before in the atomic case and

$$\hat{p}(x - R_j) = (b_j^{3/2} + a_j^2)^{1/2}. \quad (50)$$

By construction, $p_1(r) = Q_1/r$. Thus, $p(x) \leq P(x)$ if we can show that $\hat{p}(x - R_j) \leq Q_1/|x - R_j|$ for $j \neq m$. Let $|x - R_j| = u \geq r$. We require that

$$u^2[4\pi A(\gamma\lambda)^{-3/2}\delta^{3/2}(u - R)^{-6} + \delta^2(u - R)^{-8}] \leq Q_1^2. \quad (51)$$

However, the functions $u^2(u - R)^{-6}$ and $u^2(u - R)^{-8}$ are monotonically decreasing in u for $u > R$. Hence, the left-hand side of (51) is less than its value at $u = r$. But this is obviously less than $r^2 p_1(r)^2$ which is Q_1^2 . $\square$

Proposition 5. Let $0 \leq s \leq 2$, $0 \leq t \leq 2$ and let $a_1, \dots, a_K, b_1, \dots, b_K$ be $2K$ non-negative numbers. Then

$$\left[\left(\sum_{j=1}^K a_j \right)^s + \left(\sum_{j=1}^K b_j \right)^t \right]^{1/2} \leq \sum_{j=1}^K (a_j^s + b_j^t)^{1/2}. \quad (52)$$

Proof. It suffices to prove the proposition for $K = 2$ namely, for $a, A, b, B \geq 0$,

$$[(a + A)^s + (b + B)^t]^{1/2} \leq (a^s + b^t)^{1/2} + (A^s + B^t)^{1/2}. \quad (53)$$

If (53) holds then simply take $a = a_1, A = \sum_{j=2}^K a_j$ (and similarly for b, B) and use induction. Now $(a + A)^s = (a + A)^2 / (a + A)^{2-s} \leq (a^2 + 2aA + A^2) / \max(a^{2-s}, A^{2-s}) \leq a^s + 2a^{s/2}A^{s/2} + A^s$. A similar inequality holds for $(b + B)^t$. Squaring (53) and using these inequalities, it suffices to prove that

$$a^{s/2}A^{s/2} + b^{t/2}B^{t/2} \leq (a^s + b^t)^{1/2}(A^s + B^t)^{1/2}.$$

This, however follows from the Cauchy–Schwarz inequality. $\square$

3. Behaviour of N_e for small Z or small γ or large A

Although theorem 1 gives an upper bound for all values of the z_j , it is primarily useful for the large- Z behaviour of Q . In fact, the comparison function f we chose in the proof of lemma 4, (i.e. $f(x) = \delta(|x| - R)^{-4}$) may be too big when we consider small z . Since the atomic $\phi(x)$ is bounded from above by $V(x) = z|x|^{-1}$ and the function g has support on a ball of radius R and total mass 1, $\phi(x) \leq z|x|^{-1}$ for $|x| \geq R$. In particular $\phi(R) \leq zR^{-1}$, whereas the comparison function f goes to infinity at $|x| = R$. Therefore it is somewhat better to choose $f(x) = \delta(|x| - \alpha R)^{-4}$, where $\alpha = \alpha(z) = 1 - (\delta/zR^3)^{1/4}$ is such that $f(R) = zR^{-1}$. Then, proceeding as in the proof of theorem 1, one gets a z_j -dependent bound for Q . Although we do not give any details here, we point out that as z goes to zero, for an atom, this upper bound goes to $3.057A^{3/2}$ with $\gamma = \gamma_{\text{phys}}$. We know, however, that as Z goes to zero, Q vanishes because $Q < Z$ (Lieb 1981, theorem 7.23). Thus, the previous bound is not good for small Z . Getting the behaviour of Q as a function of Z for small Z is the subject of this section. The main result for Q is contained in equation (63) below.

1054 *R Benguria and E H Lieb*

We begin with a normalisation convention. Choose $z_1^0, \dots, z_K^0 > 0$ such that

$$\sum_{j=1}^K z_j^0 = 1 \quad (54)$$

and let $R_1^0, \dots, R_K^0$ be fixed, distinct points in $\mathcal{R}^3$. Let $Z > 0$ be the total nuclear charge in a molecule in which

$$z_j = Z z_j^0 \quad R_j = (A/Z) R_j^0. \quad (55)$$

In this molecule the length scale is A/Z . In TF theory, by comparison, it is $Z^{-1/3}$. As $Z \rightarrow 0$ the atoms move apart. One can also treat the case in which the R_j remain fixed as $Z \rightarrow 0$; we do not do so explicitly here, but note that in the limit $Z \rightarrow 0$ this is obviously the same as placing all the R_j at one common point.

Let us write the solution to (5) as

$$\psi(x) = Z^2 A^{-3/2} \tilde{\psi}((Z/A)x) \quad (56)$$

whence

$$\int \psi(x)^2 dx = Z \int \tilde{\psi}(x)^2 dx. \quad (57)$$

The TFW equation (5) then reads

$$(-\Delta + \tilde{\gamma} \tilde{\psi}(x)^{4/3} - \tilde{\phi}(x)) \tilde{\psi}(x) = 0 \quad (58)$$

with

$$\tilde{\gamma} = \gamma Z^{2/3} / A \quad (59)$$

$$\tilde{\phi}(x) = \tilde{V}(x) - (|x|^{-1} * \tilde{\psi}^2)(x) \quad (60)$$

$$\tilde{V}(x) = \sum_{j=1}^K z_j^0 |x - R_j^0|^{-1}. \quad (61)$$

With this scaling there is only one non-trivial parameter in the problem, $\tilde{\gamma}$. The potential $\tilde{V}$ is that of a molecule with unit total nuclear charge and $A = 1$. Our goal is to elucidate the behaviour of (58) as $\tilde{\gamma} \rightarrow 0$. From now on we shall omit the tilde on the various quantities in (58)–(61).

Formally, at least, as $\gamma \rightarrow 0$ the solution ψ_γ to (58) approaches the solution ψ_H to the Hartree equation

$$(-\Delta - \phi(x)) \psi_H = 0 \quad (62)$$

with ϕ given by (61) and (60) with ψ_H^2 . This equation (which was also used in Benguria and Lieb 1983) has a unique positive solution, ψ_H , because the proof in Benguria *et al* (1981), Lieb (1981), that (1) has a unique minimum and that this minimum is the unique solution to (5) only uses the fact that $\gamma \geq 0$. Assuming that $\int \psi^2 \rightarrow \int \psi_H^2$, (57) tells us that

$$\lim_{Z \rightarrow 0} Q/Z = \int \psi_H^2(x) dx - 1. \quad (63)$$

Proving (63) is the goal of this section.

As stated earlier, $1 < \int \psi_H^2 < 2$. For the atomic case $K = 1$, $z_1^0 = 1$, (62) has been solved numerically (Baumgartner 1983) with the result that

$$\int \psi_H(x)^2 dx = 1.21. \quad (64)$$

Thus, as $Z \rightarrow 0$, $Q \approx 0.21Z$ for an atom. The right-hand side of (63) is not known for a molecule, but we conjecture that

$$\int \psi_H(x)^2 dx \leq 1 + 0.21 K. \quad (65)$$

The point $\gamma = 0$ is not special. We shall prove the following general theorem which says that if $\gamma \rightarrow \Gamma \geq 0$ then the solution $\psi_\gamma \rightarrow \psi_\Gamma$ in a very strong sense. In particular there is strong L^2 convergence so that (63) is justified.

Theorem 6. Let ψ_γ (and $\rho_\gamma = \psi_\gamma^2$) denote the unique positive solution to the TFW equation (5) for $\gamma \geq 0$, with A fixed and with V in equation (3) fixed. (Note: condition (54) is irrelevant here.) Let $\Gamma \geq 0$ be fixed. Then, as $\gamma \rightarrow \Gamma$, $\psi_\gamma \rightarrow \psi_\Gamma$ in the following senses:

$$\nabla \psi_\gamma \rightarrow \nabla \psi_\Gamma \text{ strongly in } L^2. \quad (66)$$

$$\psi_\gamma \rightarrow \psi_\Gamma \text{ and } \rho_\gamma \rightarrow \rho_\Gamma \text{ strongly in } L^p \quad (67)$$

for all $1 \leq p \leq \infty$.

$$|x|^{-1} * \psi_\gamma^2 \rightarrow |x|^{-1} * \psi_\Gamma^2 \text{ strongly in } L^p \quad (68)$$

for all $3 < p \leq \infty$.

$$D(\psi_\gamma^2, \psi_\gamma^2) \rightarrow D(\psi_\Gamma^2, \psi_\Gamma^2) \quad (69)$$

(cf (2)).

Proof. Let $\gamma_n \geq 0$ be any sequence with $\gamma_n \rightarrow \Gamma$ and let $\psi_n \equiv \psi_{\gamma_n}$. Since ψ_Γ is unique, it suffices to show that some subsequence of ψ_n converges to ψ_Γ in the indicated senses. In the appendix it is proved that $\|\psi_\gamma\|_\infty < C_\infty = \text{constant}$, independent of γ . Since $\|\psi_\gamma\|_2^2 < 2$, then for all $2 \leq p \leq \infty$ we have $\|\psi_\gamma\|_p < C_p = \text{constant}$. With ξ_γ given by (1) and with E_γ being the minimum of ξ_γ we easily find, by considering $\xi_\gamma(\psi_\gamma^2)$ and $\xi_\Gamma(\psi_\Gamma^2)$, that

$$\lim_{\gamma \rightarrow \Gamma} E_\gamma = E_\Gamma \quad (70)$$

and also that ψ_n^2 is a minimising sequence for ξ_Γ . By the proof in Lieb (1981) and Benguria *et al* (1981) of the existence of a minimum for ξ_Γ , and the lower semicontinuity of ξ_Γ , we conclude that there is a subsequence (which we continue to denote by ψ_n) such that

$$\nabla \psi_n \rightarrow \nabla \psi_\Gamma \text{ strongly in } L^2 \quad (71)$$

$$D(\psi_n^2, \psi_n^2) \rightarrow D(\psi_\Gamma^2, \psi_\Gamma^2) \quad (72)$$

$$\psi_n \rightarrow \psi_\Gamma \text{ almost everywhere} \quad (73)$$

This proves (66) and (69). ((73) follows from the Rellich–Kondrachov theorem.)

Not only is $\psi_n(x) \leq C_\infty$ for all x , but we also have the bound (with some constants M and α)

$$\psi_n(x) \leq M \exp(-\alpha|x|^{1/2}) \quad (74)$$

for $|\gamma_n - \Gamma|$ small enough and for $|x| > R$ for some fixed R . To prove (74) we note that for some R_1 and some $Q > 0$

$$\int_B \psi_\Gamma(x)^2 dx > Z + Q/2$$

with $B = \{x \mid |x| < R_1\}$. Since $H^1(B)$ is compactly imbedded in $L^2(B)$ (Rellich-Kondrachov), there is a further subsequence such that $\psi_n \rightarrow \psi_\Gamma$ strongly in $L^2(B)$. Hence the entire sequence (see the remark at the beginning of the proof) converges strongly to ψ_Γ in $L^2(B)$. Thus,

$$\int_B \psi_n(x)^2 dx > Z + Q/4 \quad (75)$$

for n large enough. (74) follows by the proof in Benguria *et al* (1981), lemma 8 or Lieb (1981), lemma 7.24(ii). Consequently, $\psi_n(x) \leq F(x)$ for all x and n large enough, where $F(x) = C_\infty$ for $|x| \leq R$ and $F(x)$ is the right-hand side of (74) for $|x| > R$. Since $F \in L^p$ for $1 \leq p \leq \infty$, (67) follows from (73) for $1 \leq p < \infty$ by dominated convergence.

The upper bound F and (73) also imply (by dominated convergence) that the convergence in (68) is pointwise almost everywhere and that $g_\gamma \equiv |x|^{-1} * \psi_\gamma^2$ is bounded by a function of the form $G = \min(s, t/|x|)$ for suitable s and t . Again, by dominated convergence, we obtain strong convergence in (68) in L^p for $3 < p < \infty$. The L^∞ convergence in (68) follows easily from the L^2 convergence of ρ_γ to ρ_Γ together with the large $|x|$ bound (74).

To prove the L^∞ convergence in (67) note that in view of (74) it suffices to prove L^∞ convergence on bounded sets. But this follows from the fact (Benguria *et al* 1981, lemma 7, or Lieb 1981, theorem 7.9) that for any bounded set S and all $x, y \in S$, $|\psi_n(x) - \psi_n(y)| < M|x - y|^{1/2}$ for some constant M which depends on S but (it is easy to see from the proof) not on n . $\square$

4. Lower bound for the chemical potential of a neutral atom or molecule

Here, we prove a result which is somewhat related to the bound on N_c . We shall show that the chemical potential of a neutral system (K not necessarily one) is bounded from below by a constant independent of the nuclear charge. We conjecture that a similar bound from above should hold.

The chemical potential, $-\mu(N) = dE/dN$, as a function of N is nonpositive, continuous and monotonically increasing in N (for fixed nuclear charges) since $E(N)$ is convex in N in TFW theory (Lieb 1981, theorem 7.2 and theorem 7.8(iii)). Therefore the binding energy ΔE (or affinity) satisfies $|\Delta E| < \mu_0 Q$, where Q is the added charge ($Q = 1$ for an electron).

The fact that the chemical potential, $-\mu_0 = -\mu(Z)$, for a neutral atom or molecule is bounded independent of the nuclear charges agrees with what is believed to be the case for the Schrödinger equation.

Consider the TFW equation for a neutral system (which is a generalisation of (5)), i.e.

$$-A \Delta \psi + (\gamma \psi^{4/3} - \phi) \psi = -\mu_0 \psi \quad (76)$$

with ϕ given by (6), (3), or equivalently,

$$-\Delta \phi = -4\pi \psi^2 \quad x \neq R_i, \quad i = 1, 2, \dots, K. \quad (77)$$

Here $\int \psi^2 dx = Z = \sum_{i=1}^K z_i$. Our bound is the following.

Theorem 7. For a neutral system, i.e. for $\int \psi^2 dx = Z$, the chemical potential is bounded from below by,

$$-\mu_0 \geq -27\pi^2 A^2 \gamma^{-3} \quad \text{all } Z \quad (78)$$

in the units chosen in the Introduction. In particular, for the value of A chosen in Lieb (1981) to fit the Scott term in the energy, i.e. $A = 0.1859$, and with $\gamma = \gamma_{\text{phys}}$, $\mu_0 \leq 0.0105$.

Remark. Since $N_c > Z$ (Lieb 1981, theorem 7.19), μ_0 is strictly positive.

Proof. First consider the TFW equation with arbitrary $N = \int \psi^2 \leq N_c$, in which case the right side of (76) is replaced by $-\mu(N)\psi$. We know that $\mu(N) = -dE(N)/dN$ and that $\mu(N)$ is *continuous* and monotonically decreasing (Lieb 1981, theorem 7.8(iii)). Therefore (78) will be proved if we can show that for every $N > Z$, $\mu(N) < 27\pi^2 A^2 \gamma^{-3}$. This, we shall now proceed to do.

For every positive b and for all numbers $\psi \geq 0$ we have the algebraic inequality

$$b\psi^2 \leq \gamma\psi^{7/3} + d(b)\psi \quad (79)$$

with

$$d(b) = 27b^4 \gamma^{-3} / 256. \quad (80)$$

The TFW equation (with $\int \psi^2 = N > Z$) implies

$$-A \Delta \psi + b\psi^2 - \phi \psi \leq (d(b) - \mu(N))\psi.$$

Therefore, if $\mu(N) \geq d(b)$

$$-A \Delta \psi + b\psi^2 - \phi \psi \leq 0. \quad (81)$$

Now, as long as b is chosen so that $b > 4(\pi A)^{1/2}$, (77) and (81) imply that $\psi < \beta\phi$, all $x \in \mathcal{R}^3$, where β is the positive root of $b = \beta^{-1} + 4\pi\beta A$. To prove this, let $S = \{x | \psi(x) > \beta\phi(x)\}$. Obviously $R_i \notin S$. Since $\psi - \beta\phi$ is continuous in $\mathcal{R}^3 \setminus \{R_i\}$, S is open. On S ,

$$\begin{aligned} -A \Delta(\psi - \beta\phi) &\leq -b\psi^2 + \phi\psi + 4\pi A\beta\psi^2 \\ &= \beta^{-1}[\beta\phi - \beta(b - 4\pi A\beta)\psi]\psi \\ &= \beta^{-1}(\beta\phi - \psi)\psi \leq 0, \end{aligned}$$

where we have used the fact that $b - 4\pi A\beta = \beta^{-1}$. Hence $\psi - \beta\phi$ is subharmonic on S . Moreover $\psi - \beta\phi = 0$ on $\partial S \cup \{\infty\}$. Therefore S is empty and $\beta\phi(x) \geq \psi(x)$ for all $x \in \mathcal{R}^3$. Since $\psi \geq 0$, ϕ must be non-negative everywhere. On the other hand, $\phi =$

$V - |x|^{-1} * \psi^2$ and $\int \psi^2 > Z$; consequently, $\phi(x) < 0$ for sufficiently large $|x|$. This is a contradiction, and we conclude that $\mu(N) < d(b)$ whenever $b = \beta^{-1} + 4\pi\beta A$ for some $0 < \beta < \infty$. Choosing $\beta = (4\pi A)^{-1/2}$ yields the desired result, i.e. $\mu(N) < 27\pi^2 A^2 \gamma^{-3}$. $\square$

Remark. From the asymptotics of the solution of equation (76) we see that $\mu_0^{-1/2}$ somehow measures the range of the electronic density. If our conjecture is true, such a range would be independent of Z .

Acknowledgments

One of us (RB) would like to thank the Physics Department of Princeton University for their hospitality during the course of this work.

Appendix

Here, we give a bound for the L^∞ norm of the solution to the TFW equation. Such a bound is independent of γ , the constant in front of the $\psi^{7/3}$ term. This bound is used in § 3.

Lemma A.1. Let ψ be the positive solution to the TFW equation for a molecule with $V(x)$ given by (3). Then for all $\gamma > 0$

$$\|\psi\|_\infty \leq (27/16\pi)^{1/2} (Z/A)^{3/2} \|\psi\|_2$$

with $\|\psi\|_2 < (2Z)^{1/2}$ (Lieb 1981, theorem 7.23).

Proof. Because of lemma 9 in Benguria *et al* (1981), $\psi \in L^\infty$. From equation (5), $-A \Delta \psi(x) \leq V(x)\psi(x)$. First, consider a single atom with $V(x) = z|x|^{-1}$. In this case, therefore,

$$A\psi(x) \leq (4\pi)^{-1} \int z|y|^{-1}|x-y|^{-1}\psi(y) \, dy. \quad (\text{A.1})$$

Hence

$$\begin{aligned} 8\pi z^{-1} A\psi(x) &\leq \int (|y|^{-2} + |x-y|^{-2})\psi(y) \, dy \\ &= \int |y|^{-2}(\psi(y) + \psi(x-y)) \, dy. \end{aligned} \quad (\text{A.2})$$

We decompose this last integral into two terms. One integral over $\{|y| < r\}$ and the other over $\{|y| > r\}$ for any fixed $r > 0$. We have,

$$\int_{|y| < r} |y|^{-2}(\psi(y) + \psi(x-y)) \, dy \leq 8\pi r \|\psi\|_\infty \quad (\text{A.3})$$

$$\int_{|y| > r} |y|^{-2}(\psi(y) + \psi(x-y)) \, dy \leq 2\|\psi\|_2 (4\pi/r)^{1/2} \quad (\text{A.4})$$

by Hölder's inequality. Thus, substituting (A.3) and (A.4) into (A.2), we have

$$8\pi z^{-1}A\psi(x) \leq 8\pi r\|\psi\|_{\infty} + 2\|\psi\|_2(4\pi/r)^{1/2} \quad \text{all } r > 0 \quad (\text{A.5})$$

and minimising the right-hand side with respect to r we get

$$8\pi A\psi(x) \leq 6z\|\psi\|_{\infty}^{1/3}\|\psi\|_2^{2/3}(2\pi)^{2/3} \quad \text{all } x. \quad (\text{A.6})$$

In the molecular case

$$A\psi(x) \leq (4\pi)^{-1} \sum_{j=1}^K \int z_j|y - R_j|^{-1}|x - y|^{-1}\psi(y) \, dy. \quad (\text{A.7})$$

Using the same analysis (A.2)–(A.6) for each term on the right-hand side of (A.7) (but with $\{|y - R_j| \geq r_j\}$ and with r_j depending on j) we have that

$$8\pi A\psi(x) \leq 6Z\|\psi\|_{\infty}^{1/3}\|\psi\|_2^{2/3}(2\pi)^{2/3}. \quad (\text{A.8})$$

The lemma is proved by taking the supremum over x on the left-hand side of (A.8). $\square$

Remark. By making a similar decomposition, one can show that

$$\|B\psi^2\|_{\infty} \leq 3(\pi/2)^{1/3}\|\psi\|_2^{4/3}\|\psi\|_{\infty}^{2/3} \leq \|\psi\|_2^2(Z/A) \times 9 \times 2^{-5/3}. \quad (\text{A.9})$$

where

$$(B\psi^2)(x) = \int |x - y|^{-1}\psi(y)^2 \, dy$$

and $\|\psi\|_2^2 < 2Z$ (Lieb 1981, theorem 7.23). The second inequality in (A.9) comes from lemma A.1. Actually, the sharp constant in the middle term of (A.9) is $3(\pi/6)^{1/3}$, not $3(\pi/2)^{1/3}$. One can show that the maximising ρ for $\|B\rho\|_{\infty}/(\|\rho\|_1^{2/3}\|\rho\|_{\infty}^{1/3})$ is $\rho =$ characteristic function of a ball.

References

- Baumgartner B 1983 *Lett. Math. Phys.* **7** 439–41
 — 1984 *J. Phys. A: Math. Gen.* **17** 1593–602
 Benguria R, Brezis H and Lieb E H 1981 *Commun. Math. Phys.* **79** 167–80
 Benguria R and Lieb E H 1983 *Phys. Rev. Lett.* **50** 1771–4
 Lieb E H 1981 *Rev. Mod. Phys.* **53** 603–41, Errata 1982 **54** 311
 — 1982 *Commun. Math. Phys.* **85** 15–25
 — 1984a *Phys. Rev. Lett.* **52** 315–7
 — 1984b *Phys. Rev. A* **29** 3018–28
 Lieb E H and Liebman D A 1982 *Numerical Calculation of the Thomas–Fermi–von Weizsäcker Function for an Infinite Atom without Electron Repulsion* Los Alamos National Laboratory Report LA 9186-MS
 Lieb E H, Sigal I M, Simon B and Thirring W 1984 to be published
 Lieb E H and Simon B 1977 *Adv. Math.* **23** 22–116
 von Weizsäcker C F 1935 *Z. Phys.* **96** 431–58

Part V

Stability of Matter

PHYSICAL REVIEW LETTERS

VOLUME 35

15 SEPTEMBER 1975

NUMBER 11

Bound for the Kinetic Energy of Fermions Which Proves the Stability of Matter

Elliott H. Lieb*

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08540

and

Walter E. Thirring

Institut für Theoretische Physik der Universität Wien, A-1090 Wien, Austria

(Received 8 July 1975)

We first prove that $\sum |e(V)|$, the sum of the negative energies of a single particle in a potential V , is bounded above by $(4/15\pi) \int |V|^{5/2}$. This, in turn, implies a lower bound for the kinetic energy of N fermions of the form $\frac{2}{5}(3\pi/4)^{2/3} \int \rho^{5/3}$, where $\rho(x)$ is the one-particle density. From this, using the no-binding theorem of Thomas-Fermi theory, we present a short proof of the stability of matter with a reasonable constant for the bound.

The basis of all theories of bulk matter is the stability theorem of the N -electron Hamiltonian,¹

$$H_N = \sum_{i=1}^N p_i^2 - \sum_{i=1}^N \sum_{k=1}^M Z_k |x_i - R_k|^{-1} + \sum_{i>j} |x_i - x_j|^{-1} + \sum_{k>m} \frac{Z_k Z_m}{|R_k - R_m|} \geq AN. \quad (1)$$

Equation (1) has been proved by Dyson and Lenard.² Unfortunately, their analysis is complicated and their constant A gigantic, about 10^{14} . Despite subsequent improvements,³⁻⁵ a simple proof yielding a reasonable constant A has not yet been found. In this note we propose to fill this gap.

We start with the observation that if Thomas-Fermi (TF) theory were valid, then the no-binding theorem⁶ would yield the desired result because the TF energy is proportional to the number of atoms. Our goal will be to show that the TF energy, with suitably modified constants, is a lower bound to H_N . To show this, we have to demonstrate two things: (i) The TF approximation for the N -fermion kinetic energy, $K \int \rho^{5/3}$, is a lower bound for some $K > 0$; (ii) the TF approximation for the electron repulsion, $\iint \rho(x) \rho(y) |x - y|^{-1} d^3x d^3y$, can be converted into a bound by a further change of constants. The following is a sketch of our proof. Fine points of rigor, together with some variations of the inequalities given here, will be presented elsewhere.

(i) *Kinetic energy of N fermions.*—Consider the Schrödinger equation for one particle in a potential $V(x)$. Schwinger⁷ has derived an upper bound for $N_+(V)$, the number of energy levels with energies $\geq E$. For $\alpha > 0$, and with $|f(x)|_- = -f(x)$ for $f < 0$ and 0 otherwise,

$$\begin{aligned} N_{-\alpha}(V) &\leq N_{-\alpha/2}(|V + \alpha/2|_-) \leq (4\pi)^{-2} \int d^3x d^3y |V(x) + \alpha/2|_- |x - y|^{-2} \exp[-(2\alpha)^{1/2} |x - y|] |V(y) + \alpha/2|_- \\ &\leq (4\pi)^{-1} (2\alpha)^{-1/2} \int d^3x |V(x) + \alpha/2|_-^2. \end{aligned} \quad (2)$$

The last inequality is Young's. Consequently, the sum of the negative energy eigenvalues of $p^2 + V$ is

687

bounded by

$$\sum_j |e_j(V)| = \int_0^\infty N_{-\alpha}(V) d\alpha \leq (8\pi)^{-1} 2^{1/2} \int d^3x \int_0^{2|V(x)|} d\alpha \alpha^{-1/2} [V(x) + \alpha/2]^2 = (4/15\pi) \int d^3x |V(x)|_-^{5/2} \quad (3)$$

By comparison, the classical value is

$$(2\pi)^{-3} \int d^3x \int d^3p |p^2 + V(x)|_- = (15\pi^2)^{-1} \int d^3x |V(x)|_-^{5/2}. \quad (4)$$

We conjecture that (4) is actually a bound. In one dimension an analog of (3) holds with $\int |V|_-^{3/2}$, but we have a counterexample that shows that the classical value is not a bound.

Now let ψ be any N -particle normalized antisymmetric wave function of space-spin. Define

$$\rho_\pm(x) = N \int d^3x_2 \cdots d^3x_N \sum_{\sigma_2, \dots, \sigma_N} |\psi(x, x_2, \dots, x_N; \pm, \sigma_2, \dots, \sigma_N)|^2, \quad (5)$$

$$T = \langle \psi | - \sum_{i=1}^N \Delta_i | \psi \rangle, \quad (6)$$

$$K = T (\int \rho_+^{5/3} + \int \rho_-^{5/3})^{-1} > 0, \quad (7)$$

and $\pi_\pm$ are projections onto the single-particle spin states. Let $h_i = p_i^2 - (5K/3) [\rho_+^{2/3}(x_i) \pi_{i+} + \rho_-^{2/3}(x_i) \pi_{i-}]$ be a single-particle Hamiltonian and

$$H = \sum_{i=1}^N h_i.$$

If E_0 is the fermion ground-state energy of H then $E_0 \leq \langle \psi | H | \psi \rangle = T - (5K/3) (\int \rho_+^{5/3} + \int \rho_-^{5/3})$. On the other hand, E_0 is greater than or equal to the sum of all the negative eigenvalues of $p^2 - (5K/3) \rho_\pm^{2/3}$ together. By (3), $E_0 \geq - (4/15\pi) (5K/3)^{5/2} (\int \rho_+^{5/3} + \int \rho_-^{5/3})$. Combining these two inequalities^{8,9} yields

$$K \geq \frac{3}{5} (3\pi/2)^{2/3}. \quad (8)$$

With

$$\rho(x) = \rho_+(x) + \rho_-(x), \quad (9)$$

and using the convexity of $\int \rho^{5/3}$, we obtain a weakened version of (8):

$$T \geq \frac{3}{5} (3\pi/4)^{2/3} \int \rho(x)^{5/3} d^3x. \quad (10)$$

If (4) were a bound, then the TF constant, $\frac{3}{5} (3\pi^2)^{2/3}$, could be used in (10).

(ii) *Electron repulsion.*—In this paper we shall use TF theory twice; the first use is to derive a theorem about electrostatics. The TF energy functional with $\gamma > 0$ and positive charges Z_k at locations R_k is

$$\mathcal{E}_\gamma(\rho) = (3/5\gamma) \int \rho^{5/3} - \sum_{k=1}^M \int d^3x \rho(x) |x - R_k|^{-1} Z_k + \frac{1}{2} \iint \rho(x) \rho(y) |x - y|^{-1} d^3x d^3y + \sum_{j < k} Z_j Z_k |R_j - R_k|^{-1}. \quad (11)$$

For any R_j , the minimum of $\mathcal{E}_\gamma(\rho)$ occurs when $\int \rho = \sum_j Z_j$ ($j = 1$ to M), and this in turn has a minimum when the R_j are infinitely separated (no-binding theorem⁶). Thus

$$\mathcal{E}_\gamma(\rho) \geq -3.68\gamma \sum_{j=1}^M Z_j^{7/3}, \quad (12)$$

since a neutral atom of charge Z has an energy $-3.68\gamma Z^{7/3}$ in TF theory.

Consider (11) with R_j being the electron coordinates, x_j , $Z_j = 1$, $M = N$, and ρ given by (5) and (9). Multiply (11) by $|\psi|^2$ and integrate, and then use (12). Thus, for all $\gamma > 0$,

$$\langle \psi | \sum_{i < j} |x_i - x_j|^{-1} | \psi \rangle \geq \frac{1}{2} \iint \rho(x) \rho(y) |x - y|^{-1} d^3x d^3y - (3/5\gamma) \int d^3x \rho(x)^{5/3} - 3.68N. \quad (13)$$

Therefore we have an electrostatic theorem that the TF estimate for the electron repulsion [the first term on the right-hand side of (13)] is a lower bound provided one makes a kinetic energy correction and subtracts an energy proportional to the electron number.

(iii) *Stability of matter.*—Combining (10) and (13), with $\gamma > (4/3\pi)^{2/3}$ and R_k being the nuclear coordinates, yields

$$\langle \psi | H_N | \psi \rangle \geq \mathcal{E}_\delta(\rho) - 3.68N\gamma, \quad (14)$$

with $1/\delta = (3\pi/4)^{2/3} - 1/\gamma$ and ρ given by (5) and (9). A lower bound is obtained by minimizing $\mathcal{E}_\delta$ over all ρ such that $\int \rho = N$. For simplicity we shall only use the absolute minimum of $\mathcal{E}_\delta$. By (12)

$$\langle \psi | H_N | \psi \rangle \geq -3.68(N\gamma + 6 \sum_{j=1}^M Z_j^{7/3})^2. \quad (15)$$

Optimizing (15) with respect to γ yields

$$\langle \psi | H_N | \psi \rangle \geq -2.08 N \left[1 + \left(\sum_{j=1}^M \frac{Z_j^{7/3}}{N} \right)^{1/2} \right]^2. \quad (16)$$

Remarks.—(1) If the fermions are of q species (instead of 2 as in the electron case), then the right-hand side of (10) would acquire a factor $(2/q)^{2/3}$ and the right-hand side of (16) a factor $(q/2)^{2/3}$.

(2) If all $Z_j = Z$, our result (16) gives a $Z^{7/3}$ dependence instead of the known Z^2 bound.³ If $MZ \leq N$ then $MZ^{7/3}/N \leq Z^{4/3}$, which is an improvement over Ref. 3. If $MZ > N$ then we have to use the $\int \rho = N$ condition in (14). The TF no-binding theorem also holds in the subneutral case. By convexity of the TF energy in $\int \rho$, the minimum occurs for M atoms with equal electron charge N/M . If $MZ \gg N$ the energy per atom is proportional to $(N/M)^{1/3} Z^2$. Then $\langle \psi | H_N | \psi \rangle$ is bounded below by $-aN - bZ^2 M^{2/3} N^{1/3}$. While this has the correct Z dependence, it has the wrong M dependence³; $M^{2/3}$ should be replaced by $N^{2/3}$. This difficulty is inherent in TF theory. What one needs is a simple proof that if $MZ \gg N$, then one can remove most of the surplus nuclei without affecting the energy. Even for $N=1$ this is not a simple problem. Nevertheless, our present bound is proportional to the total particle number, and this is sufficient for proving the existence of the thermodynamic

limit.¹⁰

We thank J. F. Barnes and A. Martin for helpful correspondence and discussions. E. L. thanks the Institut für Theoretische Physik, Universität Wien, for its kind hospitality.

*Work supported by the U. S. National Science Foundation under Grant No. MPS 71-03375-A03 at the Massachusetts Institute of Technology.

¹The notation is $\hbar = e = 2m = 1$; x_i and p_i are electron variables; R_k and $Z_k > 0$ are nuclear coordinates and charges ($\frac{1}{4} = 1$ Ry).

²F. J. Dyson and A. Lenard, *J. Math. Phys. (N.Y.)* **8**, 423 (1967); A. Lenard and F. J. Dyson, *J. Math. Phys. (N.Y.)* **9**, 698 (1968).

³A. Lenard, in *Statistical Mechanics and Mathematical Problems*, edited by A. Lenard (Springer, Berlin, 1973).

⁴P. Federbush, *J. Math. Phys. (N.Y.)* **16**, 347, 706 (1975).

⁵J. P. Eckmann, "Sur la Stabilité de Matière" (to be published).

⁶E. Teller, *Rev. Mod. Phys.* **34**, 627 (1962). A rigorous proof of this theorem is given by E. Lieb and B. Simon, "Thomas-Fermi Theory of Atoms, Molecules, and Solids" (to be published). See also E. Lieb and B. Simon, *Phys. Rev. Lett.* **31**, 681 (1973).

⁷J. Schwinger, *Proc. Nat. Acad. Sci.* **47**, 122 (1961).

⁸The connection between $N_E(V)$ and a bound on the kinetic energy was noted by A. Martin (private communication). One can show that the converse holds; i.e., an improvement in (8) implies an improvement in (3).

⁹By numerically solving the three-dimensional variational equation for K , Eq. (7), when $N=1$, J. F. Barnes has shown that (8) holds with the TF constant $\frac{3}{5}(6\pi^2)^{2/3}$ when $N=1$ (private communication).

¹⁰J. L. Lebowitz and E. H. Lieb, *Phys. Rev. Lett.* **22**, 631 (1969); E. H. Lieb and J. Lebowitz, *Adv. Math.* **9**, 316 (1972).

which is likely to be easily satisfied.

If the decay $D^+ \rightarrow \bar{K}^0 \pi^+$ is not enhanced, as is to be expected theoretically, the upper limits on all the remaining modes, $K_s^0 \pi^+$, $\pi^+ \pi^+$, and $K^+ K^+$, lead to no constraints at all, because these modes are suppressed by $\tan^2 \theta_C$ relative to the dominant modes.

Obviously, the nonobservation of charmed hadrons at SPEAR does little to strengthen the case for the hidden-charm interpretation of the newly discovered bosons. How much the case is weakened by the new data is a topic for subjective interpretation of the bounds we have quoted above. In our minds the most damaging result is that two-body decays of D^0 account for less than 10% of its total width. While such a suppression is neither unthinkable nor unprecedented, we find it disturbing not only because it is so small but also because, if 90% of the nonleptonic decays are to three or more particles, it will be difficult to understand the observed charged-particle multiplicity. We disagree with the conclusion of Boyarski *et al.* that their upper limit on $B(D^+ \rightarrow \bar{K}^0 \pi^+ \text{ or } K^- \pi^+ \pi^+)$ violates the expectation of the conventional model by a factor of at least 3.¹² In fact, in the conventional model, with all of its pre- J/ψ baggage of sextet enhancement and 10 suppression, both decays are expected to be absent (i.e., not dominant). An incautious interpretation is that the nonobservation of these modes is good for the model, but we do not wish to go so far. Indeed, it is our feeling that if some of the upper limits, such as those given in Eqs. (2), (9), and (13), were decreased by factors of 2 or 3, the conventional charm scheme²⁻⁶ would require mod-

ification.

*Alfred P. Sloan Foundation Fellow; also at Enrico Fermi Institute, University of Chicago, Chicago, Ill. 60637.

†Operated by Universities Research Association Inc. under contract with the U. S. Energy Research and Development Administration.

¹A. M. Boyarski *et al.*, Phys. Rev. Lett. **35**, 196 (1975).

²S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D **2**, 1285 (1970).

³G. Altarelli, N. Cabibbo, and L. Maiani, Nucl. Phys. **B88**, 285 (1975). The same result was given independently by R. L. Kingsley, S. B. Treiman, F. Wilczek, and A. Zee, Phys. Rev. D **11**, 1919 (1975).

⁴Kingsley, Treiman, Wilczek, and Zee, Ref. 3.

⁵M. B. Einhorn and C. Quigg, Phys. Rev. D (to be published).

⁶M. K. Gaillard, B. W. Lee, and J. L. Rosner, Rev. Mod. Phys. **47**, 277 (1975).

⁷J.-E. Augustin *et al.*, Phys. Rev. Lett. **34**, 764 (1975).

⁸V. Chaloupka *et al.*, Phys. Lett. **50B**, 1 (1974).

⁹Y. Dothan and H. Harari, Nuovo Cimento, Suppl. No. 3, 48 (1965). The statement in Ref. 1 that the modes $K^- \pi^+ \pi^+$, $\bar{K}^0 \bar{K}^0 K^+$, $\bar{K}^0 \pi^+ \eta$, and $\bar{K}^0 \pi^+ \pi^0$ should occur in the ratios 4:4:3:1 is correct only if they are in the *totally symmetric* 10. It is not true in general.

¹⁰M. K. Gaillard and B. W. Lee, Phys. Rev. Lett. **33**, 108 (1974).

¹¹G. Altarelli and L. Maiani, Phys. Lett. **52B**, 351 (1974).

¹²The discussion surrounding Table IV of Ref. 6, on which Boyarski *et al.* apparently base their conclusion, clearly warns that these modes may be strongly suppressed.

ERRATUM

BOUND FOR THE KINETIC ENERGY OF FERMIONS WHICH PROVES THE STABILITY OF MATTER. Elliott H. Lieb and Walter E. Thirring [Phys. Rev. Lett. **35**, 687 (1975)].

Equation (2), replace $N_{-\alpha/2}(|V + \alpha/2|_-)$ with $N_{-\alpha/2}(-|V + \alpha/2|_-)$.

Equation (13), replace $-3.68N$ with $-3.68N\gamma$.

Equation (15), replace $(\dots)^2$ with $(\dots)$.

On page 687, line 15, replace $\geq E$ by $\leq E$.

Stability of Coulomb Systems with Magnetic Fields

I. The One-Electron Atom

Jürg Fröhlich^{1,*}, Elliott H. Lieb^{2,**}, and Michael Loss^{2,***}

¹ Theoretical Physics, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

² Departments of Mathematics and Physics, Princeton University, Jadwin Hall, P.O.B. 708, Princeton, NJ 08544 USA

Abstract. The ground state energy of an atom in the presence of an external magnetic field B (with the electron spin-field interaction included) can be arbitrarily negative when B is arbitrarily large. We inquire whether stability can be restored by adding the self energy of the field, $\int B^2$. For a hydrogenic like atom we prove that there is a critical nuclear charge, z_c , such that the atom is stable for $z < z_c$ and unstable for $z > z_c$.

1. Introduction

The problem of the stability of an atom (i.e. the finiteness of its ground state energy) was solved by the introduction of the Schrödinger equation in 1926. While it is true that Schrödinger mechanics nicely takes care of the $-ze^2/r$ Coulomb singularity at $r=0$ (here $z|e|$ is the nuclear charge), a more subtle problem that has to be considered is the interaction of the atom with an external magnetic field $B(x)$ with vector potential $A(x)$ and $B = \text{curl } A$. In this paper the problem of the one-electron atom in a magnetic field is studied; in a subsequent paper [6] some aspects of the many-electron and many-nucleus problem will be addressed.

Units. Our unit of length will be half the Bohr radius, namely $l = \hbar^2/(2me^2)$. The unit of energy will be 4 Rydbergs, namely $2me^4/\hbar^2 = 2mc^2\alpha^2$, where α is the fine structure constant $e^2/(\hbar c)$. The magnetic field B is in units of $|e|/(l^2\alpha)$. The vector potential satisfies $B = \text{curl } A$. The magnetic field energy ($\int B^2/8\pi$) is, in these units,

$$\varepsilon \int B^2, \quad 1/\varepsilon = 8\pi\alpha^2. \quad (1.1)$$

* Work partially supported by U.S. National Science Foundation grant DMS-8405264 during the author's stay at the Institute for Advanced Study, Princeton, NJ, USA

** Work partially supported by U.S. National Science Foundation grant PHY-8116101-A03

*** Work partially supported by U.S. and Swiss National Science Foundation Cooperative Science Program INT-8503858. Current address: Institut f. Mathematik, FU Berlin, Arnimallee 3, D-1000 Berlin 33

The first problem to be considered is one in which the electron spin is neglected. The Hamiltonian in this case is

$$H' = (p - A)^2 - z/|x| \quad (1.2)$$

(with $p = i\nabla$). H' presents no interesting problem as far as stability is concerned because the effect of including A is *always* to raise the ground state energy. (Reason: For any ψ , $(\psi, (p - A)^2 \psi) \geq (|\psi|, p^2 |\psi|)$. This is essentially Kato's inequality [4] (see e.g. [13]). On the other hand $(\psi, |x|^{-1} \psi) = (|\psi|, |x|^{-1} |\psi|)$, so we can lower the energy by replacing ψ by $|\psi|$ and setting $A = 0$.)

The problem becomes interesting when the electron spin is included, and this problem is the subject of this paper. The wave function ψ is a two-component (complex valued) spinor:

$$\psi(x) = (\psi_1(x), \psi_2(x)). \quad (1.3)$$

The Hamiltonian is

$$H = (p - A)^2 - \sigma \cdot B(x) - z/|x| \quad (1.4)$$

$$= [\sigma \cdot (p - A)]^2 - z/|x| \quad (1.5)$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices. The first term in (1.5) is the Pauli kinetic energy and is the non-relativistic approximation to the Dirac operator.

The ground state energy $E_0(B, z)$ of H is always finite but depends on B in such a way that $E_0 \rightarrow -\infty$ as $B \rightarrow \infty$ (for a constant field), roughly as $-(\ln B)^2$, see [1]. What prevents B from spontaneously growing large and driving E_0 towards $-\infty$? (We do not inquire into the source of this B , but simply assume that nature will always contrive to lower the energy, if possible.) The answer, which we shall take as a hypothesis here, is that the price to be paid is the field energy $\int B^2/8\pi$. Thus, we are led to consider (in our units) $H + \varepsilon \int B(x)^2 dx$ and ask whether

$$E(B, z) = E_0(B, z) + \varepsilon \int B^2 \quad (1.6)$$

is bounded below *independent* of B . This problem is important in the analysis of stability in non-relativistic quantum electrodynamics [we have omitted the term $\int E^2$ which makes (1.6) a lower bound and which makes the magnetic field classical]. We define

$$E(z) = \inf_B E(B, z). \quad (1.7)$$

In the remainder of this introduction we shall first outline our results about $E(z)$, then discuss their physical interpretation and finally formulate some preliminary mathematical facts and notation.

We show that there is a critical value of z (called z_c) such that

$$\begin{aligned} E(z) &= -\infty & \text{for } z > z_c, \\ E(z) &\text{ is finite } & \text{for } z < z_c. \end{aligned} \quad (1.8)$$

The value of z_c is proportional to $1/\alpha^2$ (not $1/\alpha$ as in the case of the Dirac equation or the “relativistic” Schrödinger equation [3]). Section II (in conjunction with [8]) contains the proof that z_c is finite. The fact that $z_c \neq \infty$ is intimately connected with

the fact that the equation

$$\sigma \cdot (p - A)\psi = 0 \quad (1.9)$$

has a non-zero solution with $\psi \in H^1$ and $A \in L^6$. When we first worked on this problem we realized this connection and proved (Sect. II) that

$$z_c = \inf \int B^2 / (\psi, |x|^{-1} \psi), \quad (1.10)$$

where the infimum is over all solutions to (1.9). (Clearly, any solution to (1.9) has zero kinetic energy so that if z exceeds the right side of (1.10) the total energy can be driven to $-\infty$ by the scaling $\psi(x) \rightarrow \lambda^{3/2} \psi(\lambda x)$, $A(x) \rightarrow \lambda A(\lambda x)$. The converse is the difficult part of Sect. II.) At first it was unknown whether or not (1.9) has a solution, but now several have been found [8].

Section III gives a lower bound to z_c (which we call z_c^L):

$$z_c > z_c^L = (24.0)/(8\pi\alpha^2) > 17,900. \quad (1.11)$$

This is far better than that needed for physics. For all $z < z_c^L$ we also derive a lower bound for $E(z)$:

$$E(z) \geq -\frac{1}{4}z^2 - z^3(32z_c^L)^{-1}(1 - \frac{3}{4}z/z_c^L)^{3/2}. \quad (1.12)$$

[Note that $E(z)$ is trivially less than $-\frac{1}{4}z^2$, which is the ground state energy for $B \equiv 0$.]

The B field that causes $E(z)$ to diverge when $z > z_c$ is highly inhomogeneous (both in magnitude and direction) near the nucleus. In astrophysical and other applications [9, 11] one is interested in studying atoms and ions in very strong, external magnetic fields with the property that the *direction of the magnetic field is constant* over distance scales many times the scales of atomic physics, to a very good approximation. Theoretical astrophysicists have carried out large-scale numerical calculations of the spectra of atoms and ions in very strong magnetic fields and have tried to correlate theoretical predictions with experimental data. As a modest contribution to the mathematical foundations of this kind of work, we establish stability of one-electron atoms in *arbitrarily strong* magnetic fields whose direction (but *not* magnitude) is *constant* in a neighborhood of the atom. This is done in Sect. IV, where we prove that $E(z)$ is always finite in this case. An open problem for further investigation is the analysis of $E_0(B, z)$ for magnetic fields that are curl free in a neighborhood of the atom.

Before proceeding to the physical interpretation, we note in passing that the electron g factor was taken to be 2 in (1.4). If we replace the $\sigma \cdot B$ term in (1.4) by $\frac{1}{2}g\sigma \cdot B$ then two cases arise:

$g < 2$. Here we can write the kinetic energy as $\frac{1}{2}g[\sigma \cdot (p - A)]^2 + (1 - \frac{1}{2}g)(p - A)^2$. The first term is nonnegative and the second, when combined with $-z/|x|$ gives a Hamiltonian of type H' in (1.2). This is bounded below by $-\frac{1}{2}z^2/(2 - g)$, and hence $E(z)$ is always finite.

$g > 2$. Here, $E(z) = -\infty$ for all z , *including* $z = 0$. To see this, let B be a field which is constant $= B(0, 0, 1)$ over a large cube of length L , with $A = \frac{1}{2}B(x_2, -x_1, 0)$ inside this cube. Let B drop to zero outside the cube so that $I = \int B^2 < \infty$. Take ψ to be a ground state Landau orbital (cut off in the x_3 direction so that $\psi \in L^2$), i.e.

$$\psi(x) = (\text{const})(1, 0) \exp[-\frac{1}{4}B(x_1^2 + x_2^2)] \cos(\pi x_3/L)$$

and $\psi(x)=0$ for $|x_3|>L/2$. With B fixed and with L big enough, we can have $(\psi, [\sigma \cdot (p-A)]^2 \psi) \leq \frac{1}{4}(g/2-1)B$ and $(\psi, \sigma \cdot B \psi) \geq \frac{1}{2}B$. Also $I \leq 2B^2 L^3$. The total energy (with $z=0$) is less than $-\frac{1}{4}(g/2-1)B + 2B^2 L^3 \varepsilon$. Now, let $\lambda > 0$ and replace $\psi(x)$ by $\lambda^{3/2} \psi(\lambda x)$, $A(x)$ by $\lambda A(\lambda x)$ and $B(x)$ by $\lambda^2 B(\lambda x)$. The energy is then less than $-\frac{1}{4}\lambda^2(g/2-1)B + 2\lambda B^2 L^3 \varepsilon$. (This scaling is exact and will be employed frequently in the sequel.) As $\lambda \rightarrow \infty$, the energy tends to $-\infty$, so stability never holds.

Since physically $g > 2$ because of Quantum Electrodynamics (QED) effects, it is clear that if we try to “improve” (1.4) by replacing $\sigma \cdot B$ by $\frac{1}{2}g\sigma \cdot B$ we shall get an inconsistent theory. The only truly consistent procedure is to include *all* QED effects, and this is outside the scope of this paper.

The foregoing aside about the g -factor leads us to the question of the physical content of the results of this paper, (1.8)–(1.12). There are two ways to view them. The first is to observe that (1.8) and (1.11) show that atomic physics with the Hamiltonian (1.4) contains no seeds of instability for small z (small meaning $z < 17,900$) and that perturbation theory (in B) can be safely employed for very small B . (Of course one should also analyze the many-electron and many-nucleus problem to be certain about this conclusion. We are unable to do this fully, but in a subsequent paper [6] we do successfully analyze two problems: the one-electron, many-nucleus problem and the one-nucleus, many-electron problem, i.e. the full atom.) The fact that the theory is well behaved for small z is not entirely a trivial matter, especially when the situation is contrasted with that for spin-spin interactions (either electron-electron or electron-nucleus). Here, one adds a two-body term $\sigma^a \cdot \sigma^b |x|^{-3} - 3(\sigma^a \cdot x)(\sigma^b \cdot x)|x|^{-5}$, where x is the vector between particles a and b . The $|x|^{-3}$ singularity is not integrable and, in particular it cannot be controlled by the kinetic energy. Thus, a system with this interaction is *always* unstable in our sense. The treatment of the interaction by perturbation theory, is not really a consistent procedure.

Of course, it is always possible to restore stability by cutting off the Coulomb or spin-spin interactions at the Compton wavelength of the electron, but then the theory would depend critically on this wavelength. Stability, in the sense we use it, implies that the Schrödinger equation for electrons and nuclei is independent of the electron’s Compton wavelength—in conformity with what is always assumed to be the case.

The second viewpoint is to emphasize the breakdown of (1.4) when $z > z_c$ and to say that magnetic interactions impose an upper bound on $z\alpha^2$. Here we are treading on shaky ground. If we specify ψ and ask what B minimizes $(\psi, H\psi) + \varepsilon \int B^2$, we easily find that Maxwell’s equation takes the form

$$2\varepsilon \operatorname{curl} B(x) = j(x) = 2 \operatorname{Re} \langle \psi, (p-A)\psi \rangle(x) + \operatorname{curl} \langle \psi, \sigma \psi \rangle(x). \quad (1.13)$$

[Notation. $(\psi, H\psi)$ has been used to denote the usual expectation, including the x -integration. $\langle \psi, \sigma \psi \rangle(x)$ denotes the inner product with respect to the spinor indices *only*, and hence it is a function of x . $\langle \psi, \psi \rangle(x) \equiv |\psi(x)|^2 = |\psi_1(x)|^2 + |\psi_2(x)|^2$.]

The first term in j is the electron current ($p-A$ is the velocity). The second term is the “spin” current; it is conserved. The B field in (1.13) cannot be viewed as external; it is, in fact, generated by the electron as (1.13) shows. It is this B field that causes the breakdown when $z > z_c$. [Technical note. In Sect. II we choose a special

pair ψ, A with $\sigma \cdot (p - A)\psi \equiv 0$, so that the right side of (1.13) is zero. For this ψ , the B field we use is not exactly optimal [because (1.13) is not satisfied], but the error becomes inconsequential when we employ the λ -scaling $\psi(x) \rightarrow \lambda^{3/2}\psi(\lambda x)$ and $A(x) \rightarrow \lambda A(\lambda x)$.]

The instability of (1.6) for $z > z_c$ might indicate a qualitative change in the behaviour of non-relativistic quantum electrodynamics (QED), e.g. some kind of phase transition or an intrinsic instability, as z becomes large. For a compelling argument in this direction we would, however, have to include the term $\int \mathbf{E}^2$ in the Hamiltonian, quantize the electromagnetic field and properly renormalize the theory. Our calculations can be viewed as a quasi-classical approximation to that theory. The fact that this approximation exhibits an instability, for large z , should, by experience, be seen as a warning that the full theory might also exhibit a drastic change in behaviour, for large z .

Physically, our instability result for $z > z_c$ is, of course, quite irrelevant, because $z_c > 17,000$. Nuclei with nuclear charge above ~ 100 are not known to exist in nature, and even if nuclei with $z \sim 10,000$ existed electrons moving in their field would be highly relativistic particles, so that our use of non-relativistic kinematics is not justified for values of z where the instability occurs. Nevertheless, we feel that it is an interesting mathematical problem to explore the consistency of this model even beyond the domain, where the approximation is justified.

As remarked after (1.12), the interaction given in (1.4) *lowers* the energy. In contrast to this, the Lamb shift, which is obtained from a proper QED calculation (but only in perturbation theory), is a *raising* of the energy. Furthermore the Lamb shift is of order $z^4\alpha^3$ (apart from logarithmic corrections) which contrasts with our lowering (1.12) which is of order $z^3\alpha^2$. Our result is not directly comparable with the Lamb shift since the latter requires a fully quantized theory with renormalization.

Now we turn to the mathematical preliminaries to the rest of this paper. Some notation will be introduced and, more importantly, a careful discussion of the class of functions (A, B, ψ) will be given.

First, consider the B field. In order that (1.1) make sense we obviously require $B \in L^2(\mathbb{R}^3)$. [Notation. For vector fields $(A$ or $B)$

$$\|A\|_p \equiv \|(A \cdot A)^{1/2}\|_p, \quad (1.14)$$

where $A = (A_1, A_2, A_3)$ and $A \cdot A = \sum |A_i|^2$. For spinors ψ

$$\|\psi\|_p = \|\langle \psi, \psi \rangle^{1/2}\|_p, \quad (1.15)$$

where $\langle \psi, \psi \rangle(x) = |\psi_1(x)|^2 + |\psi_2(x)|^2 = \left\{ \sum_i |\langle \psi, \sigma_i \psi \rangle(x)|^2 \right\}^{1/2}$. For gradients

$$\|\nabla A\|_2 = \left\| \left(\sum_{i,j} |\partial_i A_j|^2 \right)^{1/2} \right\|_2 = \left\{ \sum_{i,j} \int |\partial_i A_j|^2 \right\}^{1/2} \quad (1.16)$$

with $\partial_i = \partial/\partial x_i$, $i = 1, 2, 3$. A similar formula holds for $\|\nabla \psi\|_2$.] The vector potential, A , satisfies $\text{curl } A = B$, but A is determined only up to a gauge (i.e. $A \rightarrow A + \nabla \Phi$). Gauge transformations on ψ (i.e. $\psi \rightarrow e^{i\Phi}\psi$) can be nasty ($e^{i\Phi}$ can have very bad differentiability properties). This problem is avoided by fixing a gauge, namely the Coulomb gauge, $\text{div } A = 0$. Additionally, it will be convenient to have the

formal identity (when $\operatorname{div} A = 0$)

$$\|B\|_2^2 = \int B^2 = \|\nabla A\|_2^2. \quad (1.17)$$

The danger is that A might conceivably have bad decay properties at infinity which would prevent the necessary integrations by parts in (1.17). This problem is resolved in Appendix A. Notice that if $\nabla A \in L^2$ and if $A(x) \rightarrow 0$ as $|x| \rightarrow \infty$ in a weak sense, see [2], then, by the Sobolev inequality

$$\|\nabla A\|_2 \geq S \|A\|_6, \quad (1.18)$$

so A is automatically in L^6 . Theorem A.1 in the appendix states that when $B \in L^2$ there is a *unique* A satisfying

$$\operatorname{curl} A = B, \quad \operatorname{div} A = 0 \text{ in } \mathcal{D}', \text{ and } A \in L^6, \quad (1.19)$$

and this A also satisfies (1.17), which implies $\nabla A \in L^2$. Here, $\mathcal{D}'$ denotes the usual space of distributions. This is the A we shall use (except in Theorems 2.1 and A.2 where only the assumption $A \in L^6$ is used).

Next we turn to the spinor field ψ which obviously must be in L^2 . To avoid operator domain questions we shall interpret the first term in (1.5) as a quadratic form $Q = \|\sigma \cdot (p - A)\psi\|_2^2$. Theorem A.2 states that if $\sigma \cdot (p - A)\psi \in L^2$, $\psi \in L^2$, and $A \in L^6$, then automatically $\nabla \psi \in L^2$. This, in turn, implies that $(\psi, |x|^{-1}\psi) < \infty$ by (1.18), or by the well known uncertainty principle for the hydrogen atom,

$$\|\psi\|_2 \|\nabla \psi\|_2 \geq (\psi, |x|^{-1}\psi). \quad (1.20)$$

Therefore, we introduce the class of function pairs

$$\mathcal{C} = \{\psi, A | \psi \in H^1(\mathbb{R}^3), \|\psi\|_2 = 1, A \in L^6(\mathbb{R}^3), \operatorname{div} A = 0, \nabla A \in L^2(\mathbb{R}^3)\}. \quad (1.21)$$

[$\psi \in H^1$ means $\psi \in L^2$ and $\nabla \psi \in L^2$. The set of functions f satisfying $f \in L^6(\mathbb{R}^3)$, $\nabla f \in L^2(\mathbb{R}^3)$ is sometimes called $D^{1,2}(\mathbb{R}^3)$; it is the completion of $H^1(\mathbb{R}^3)$, not in the H^1 -norm ($\|f\|_2^2 + \|\nabla f\|_2^2$)^{1/2}, but in the norm $\|\nabla f\|_2$.]

For functions in $\mathcal{C}$ the following energy functional is a generalization of $(\psi, H\psi) + \varepsilon \int B^2$.

$$\mathcal{E}(\psi, A) = \|\sigma \cdot (p - A)\psi\|_2^2 + \varepsilon \|B\|_2^2 - z(\psi, |x|^{-1}\psi), \quad (1.22)$$

and each term in (1.22) is well defined. The ground state energy is

$$E(z) = \inf\{\mathcal{E}(\psi, A) | (\psi, A) \in \mathcal{C}\}. \quad (1.23)$$

Theorem 2.4 states that when $E(z)$ is finite, the infimum in (1.23) is a minimum.

Another class we shall need is

$$\mathcal{F} = \{\psi, A | (\psi, A) \in \mathcal{C} \text{ and } \sigma \cdot (p - A)\psi = 0\}. \quad (1.24)$$

Notice that when $(\psi, A) \in \mathcal{C}$, then each term $p\psi$ and $A\psi$ makes sense as L^2 function. In Sect. II the formula

$$z_c = \varepsilon \inf\{\|B\|_2^2 / (\psi, |x|^{-1}\psi) | (\psi, A) \in \mathcal{F}\} \quad (1.25)$$

will be derived. Theorem 2.5 states that the infimum in (1.25) is actually a minimum.

II. A Basic Theorem and a Formula for z_c

Heuristically, if $\mathcal{E}(\psi, A)$ is unbounded for a certain z , we expect ψ and A to blow up in some sense. The following theorem is essential for understanding this blowup. It is stated in general terms, but its use for our problem will be clarified shortly; it will yield a formula for z_c .

Theorem 2.1. *Let ψ_n be a sequence of spinor valued functions on $\mathbb{R}^3$ and A_n a sequence of vector fields on $\mathbb{R}^3$ satisfying (for some fixed $3 < p < \infty$)*

- (i) $d_1 \leq \|\psi_n\|_2 \leq d_2$ for some constants $d_2 \geq d_1 > 0$.
- (ii) $\|\nabla \psi_n\|_2 \rightarrow \infty$ as $n \rightarrow \infty$.
- (iii) $\|A_n\|_p \leq D \|\nabla \psi_n\|_2^s$ for some $D > 0$, where $s = 1 - \frac{3}{p}$.
- (iv) $\|\sigma \cdot (p - A_n) \psi_n\|_2 \leq C_n \|\nabla \psi_n\|_2$ for some sequence $\{C_n\}_{n=1}^\infty$ with $C_n \rightarrow 0$ as $n \rightarrow \infty$. Define $1/\lambda_n = \|\nabla \psi_n\|_2$ (whence $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$), $\phi_n(x) = \lambda_n^{3/2} \psi_n(\lambda_n x)$ and $\alpha_n(x) = \lambda_n A_n(\lambda_n x)$. Then
 - (a) $\liminf_{n \rightarrow \infty} \|\alpha_n\|_p \geq c > 0$.
 - (b) There exists a subsequence (which we continue to denote by n) and functions ϕ and α , and a sequence of points $x_n \in \mathbb{R}^3$ such that $\tilde{\phi}_n(x) \equiv \phi_n(x - x_n) \rightharpoonup \phi(x) \neq 0$ weakly in $H^1(\mathbb{R}^3)$, $\tilde{\alpha}_n(x) \equiv \alpha_n(x - x_n) \rightharpoonup \alpha(x) \neq 0$ weakly in $L^p(\mathbb{R}^3)$. Moreover,

$$\sigma \cdot (p - \alpha) \phi = 0. \quad (2.1)$$
 - (c) If the original sequence has the property that ϕ_n does not converge weakly to zero in $H^1(\mathbb{R}^3)$, then the statement in part (b) holds with $x_n \equiv 0$.

Proof. Clearly by (i) and the definition of ϕ_n we have that ϕ_n is uniformly bounded in $H^1(\mathbb{R}^3)$. By (iii)

$$\|\alpha_n\|_p = \lambda_n^{\frac{p-3}{p}} \|A_n\|_p \leq D \lambda_n^{1-s-3/p} = D, \quad (2.2)$$

so α_n is uniformly bounded in $L^p(\mathbb{R}^3)$. By (iv) $\|\sigma \cdot (p - \alpha_n) \phi_n\|_2 \leq C_n \rightarrow 0$ as $n \rightarrow \infty$. By the triangle inequality and Hölder's inequality with $q = 2p/(p-2)$ we have

$$C_n \geq \|\sigma \cdot (p - \alpha_n) \phi_n\|_2 \geq \|\nabla \phi_n\|_2 - \|\alpha_n \phi_n\|_2 \geq 1 - \|\alpha_n\|_p \|\phi_n\|_q. \quad (2.3)$$

Note that $\|(\sigma \cdot p) \phi\|_2 = \|\nabla \phi\|_2$ and that $\|(\sigma \cdot \alpha) \phi\|_2 = \|\alpha \phi\|_2$. The latter uses the trivial identity $(\sigma \cdot \alpha)^2 = \alpha^2$. The former uses the same identity in Fourier space $(\sigma \cdot p)^2 = p^2$, and this is justified since $\phi \in H^1$. Also note that $2 < q < 6$. Inequality (2.3) will be used in two ways. Since ϕ_n is uniformly bounded in H^1 we have, by Sobolev's inequality, that $\|\phi_n\|_q \leq d_q$, and hence $\|\alpha_n\|_p \geq (1 - C_n)/d_q$, which proves (a). On the other hand using (2.2) together with (2.3) we find $\|\phi_n\|_q \geq (1 - C_n)/D$, and hence $\liminf \|\phi_n\|_q \geq 1/D > 0$. Since $\|\phi_n\|_2 \leq d_2$ and $\|\phi_n\|_6 \leq d_6$, Lemma 2.1 and Lemma 2.2 below prove the existence of a sequence $x_n \in \mathbb{R}^3$ and a subsequence ϕ_n such that $\tilde{\phi}_n \rightharpoonup \phi \neq 0$ weakly in $H^1(\mathbb{R}^3)$. By passing, if necessary, to a further subsequence we can assume that $\tilde{\alpha}_n \rightharpoonup \alpha$ weakly in $L^p(\mathbb{R}^3)$. Next we show that $\tilde{\phi}_n \tilde{\alpha}_n \rightharpoonup \phi \alpha$ componentwise weakly in $L^2(\mathbb{R}^3)$. To show that $f_n \rightharpoonup f$ in $L^2(\mathbb{R}^3)$, it suffices to prove that $f_n \rightharpoonup f$ in $L^2(K)$ for every compact $K \subset \mathbb{R}^3$. By the Rellich-Kondrachov theorem, $\tilde{\phi}_n \rightarrow \phi$ strongly in $L^q(K)$ (since $2 < q < 6$), whence $\tilde{\phi}_n \tilde{\alpha}_n \rightharpoonup \phi \alpha$ in $L^2(K)$. Thus, we have proved that $g_n \equiv \sigma \cdot (p - \tilde{\alpha}_n) \tilde{\phi}_n \rightharpoonup \sigma \cdot (p - \alpha) \phi \equiv g$ weakly in $L^2(\mathbb{R}^3; \mathbb{C}^2)$. But we already noted that $\|g_n\|_2 \rightarrow 0$, which implies (by the weak lower

semicontinuity of the norm) that $g=0$. Obviously, $\alpha \neq 0$ because otherwise we should have $(\sigma \cdot p)\phi=0$ with $\phi \neq 0$, which is impossible [recall that $\|(\sigma \cdot p)\phi\|_2 = \|\nabla \phi\|_2$]. This proves (b).

To prove (c) we note a trivial generalization of the Banach-Alaoglu theorem: Since $\phi_n \rightharpoonup 0$ we can find a subsequence such that $\phi_n \rightarrow \phi$ and $\phi \neq 0$. The rest of the proof is the same as in (b), except that Lemma 2.2 is not needed. $\square$

Lemma 2.1. *Let g be a measurable function on a measure space such that for $p < q < r$ fixed and for some C_p, C_q, C_r all > 0 ,*

- (i) $\|g\|_p^p \leq C_p$,
- (ii) $\|g\|_r^r \leq C_r$,
- (iii) $\|g\|_q^q \geq C_q > 0$.

Then $f(\varepsilon) \equiv \text{meas}\{x | |g(x)| \geq \varepsilon\} > C$ for some fixed $\varepsilon, C > 0$ depending on p, q, r, C_p, C_q, C_r , but not on g .

Proof. From the fact that $f(\varepsilon)$ is monotone non-increasing and that $\int g^p = p \int_0^\infty f(\varepsilon) \varepsilon^{p-1} d\varepsilon$, we have $C_p \geq p \int_0^R \varepsilon^{p-1} f(\varepsilon) d\varepsilon \geq R^p f(R)$ or

$$f(\varepsilon) \leq \varepsilon^{-p} C_p, \quad \text{all } \varepsilon > 0. \quad (2.4)$$

Similarly,

$$f(\varepsilon) \leq \varepsilon^{-r} C_r, \quad \text{all } \varepsilon > 0. \quad (2.5)$$

Define S and T by

$$\begin{aligned} qC_p S^{q-p} &= \frac{1}{4}(q-p)C_q, \\ qC_r T^{q-r} &= \frac{1}{4}(r-q)C_q. \end{aligned}$$

From (2.4)

$$q \int_0^S f(\varepsilon) \varepsilon^{q-1} d\varepsilon \leq qC_p \int_0^S \varepsilon^{q-p-1} d\varepsilon = \frac{1}{4}C_q. \quad (2.6)$$

Similarly, from (2.5)

$$q \int_T^\infty f(\varepsilon) \varepsilon^{q-1} d\varepsilon \leq \frac{1}{4}C_q, \quad (2.7)$$

(2.6) and (2.7) imply that $S < T$ and that

$$I \equiv q \int_S^T f(\varepsilon) \varepsilon^{q-1} d\varepsilon \geq \frac{1}{2}C_q.$$

But $I \leq f(S)(T^q - S^q)$ since f is monotone nonincreasing. This proves the lemma (with $\varepsilon \equiv S$) since S and T are explicitly given independent of f . $\square$

Lemma 2.2 [5]. *Let $1 < p < \infty$ and let $\{f_n\}_{n=1}^\infty$, be a uniformly bounded sequence of functions in $W^{1,p}(\mathbb{R}^d)$ with the property that the Lebesgue measure of $\{x | |f_n(x)| > \varepsilon\} > C$ for some fixed constants C and $\varepsilon > 0$. Then there exists a sequence of translations $\{\tau_n\}_{n=1}^\infty$ of $\mathbb{R}^d$, $\tau_n y = y + x_n$, $F_n(y) \equiv f_n(\tau_n y) = f_n(y + x_n)$, such that, for some subsequence, $F_n \rightharpoonup F$ weakly in $W^{1,p}$ and $F \neq 0$.*

Remark. The proof in [5] was given for real valued functions. It is easy to see that the lemma holds for complex valued functions by considering separately real and imaginary parts. The same argument then carries over to complex spinors. We recall that $W^{1,p}$ consists of all functions in L^p whose first derivatives are in L^p . Note that $W^{1,2} = H^1$.

Let us now apply Theorem 2.1 to the proof of formula (1.10) for z_c . Suppose that z is such that $E = -\infty$. This means there exists a sequence of pairs $(\psi_n, A_n) \in \mathcal{C}$ such that as $n \rightarrow \infty$,

$$E_n = \mathcal{E}(\psi_n, A_n) = \|\sigma \cdot (p - A_n)\psi_n\|_2^2 - z(\psi_n, |x|^{-1}\psi_n) + \varepsilon \int B_n^2 dx \quad (2.8)$$

tends to $-\infty$. We verify the assumptions of Theorem 2.1 for the sequence (ψ_n, A_n) . (The usage of ϕ_n, α_n as in Theorem 2.1 will be continued.) (i) is trivial since $\|\psi_n\|_2 = 1$.

Observe that $-z(\psi_n, |x|^{-1}\psi_n)$ is the only negative term in (2.8) and hence $(\psi_n, |x|^{-1}\psi_n) \rightarrow \infty$ as $n \rightarrow \infty$. (ii) follows from the inequality $(\psi_n, |x|^{-1}\psi_n) \leq \|\nabla \psi_n\|_2 \|\psi_n\|_2$.

We can choose $E_n < 0$ (all n) and we find

$$\|\sigma(p - A_n)\psi_n\|_2^2 + \varepsilon \int B_n^2 \leq z(\psi_n, |x|^{-1}\psi_n) \leq z \|\nabla \psi_n\|_2. \quad (2.9)$$

(iv) holds with $C_n = z^{1/2} \|\nabla \psi_n\|_2^{-1/2}$.

From (2.9) we also obtain $\|B_n\|_2^2 \leq (z/\varepsilon) \|\nabla \psi_n\|_2$. On the other hand, Sobolev's inequality gives

$$\|B_n\|_2^2 = \sum_i \|\nabla A_{i,n}\|_2^2 \geq S^2 \|A_n\|_6^2. \quad (2.10)$$

Thus, (iii) holds with $p=6$ and $s=\frac{1}{2}$.

The conclusions (a) and (b) of Theorem 2.1 thus hold for the sequence (ψ_n, A_n) . It is easily seen that conclusion (c) also holds, for suppose $\phi_n \rightarrow 0$ weakly in $H^1(\mathbb{R}^3, \mathbb{C}^2)$. This would imply that $b_n \equiv (\phi_n, |x|^{-1}\phi_n) \rightarrow 0$ as $n \rightarrow \infty$. [To prove this, let B_R be the ball of radius R centered at 0 and χ_R its characteristic function. Note that, by Rellich-Kondrachov, $\phi_n \rightarrow 0$ strongly in $L^4(B_R)$. Then, writing $b_n = b_n^+ + b_n^-$, with $b_n^- = (\phi_n, |x|^{-1}\chi_R\phi_n)$, we have that $b_n^- \rightarrow 0$. However $b_n^+ \leq R^{-1}$ since $\|\phi_n\|_2 = 1$. Then let $R \rightarrow \infty$.] The energy can be written [using λ_n of Theorem 2.1, $\beta_n(x) \equiv \lambda_n^2 B_n(\lambda_n x)$] as

$$E_n = \mathcal{E}(\psi_n, A_n) = \lambda_n^{-1} \{ \lambda_n^{-1} \|\sigma \cdot (p - \alpha_n)\phi_n\|_2^2 + \varepsilon \|\beta_n\|_2^2 - z b_n \}. \quad (2.11)$$

If $b_n \rightarrow 0$ then, since $E_n < 0$, $\beta_n \rightarrow 0$ strongly in L^2 . By (2.10), $\alpha_n \rightarrow 0$ strongly in L^6 , which contradicts conclusion (a) of Theorem 2.1.

In the foregoing we did not actually use the fact that $E_n \rightarrow -\infty$, but only the facts that $E_n < 0$ and that the Coulomb energy diverges. The foregoing analysis was actually the proof of the following

Theorem 2.2. Let $(\psi_n, A_n) \in \mathcal{C}$ be a sequence satisfying

$$E_n < 0 \quad \text{and} \quad \limsup_{n \rightarrow \infty} (\psi_n, |x|^{-1}\psi_n) = \infty.$$

Then conclusions (a) and (c) of Theorem 2.1 hold for this sequence. Moreover, for the subsequence given by Theorem 2.1 (c),

$$(\phi_n, |x|^{-1}\phi_n) \rightarrow (\phi, |x|^{-1}\phi) \neq 0.$$

Remark. For any minimizing sequence (for any z) we can always assume $E_n < 0$, since one can always take the pair $\psi =$ ground state hydrogenic function and $A \equiv 0$.

Let us define

$$\hat{z} = \inf_{\mathcal{F}} \varepsilon \|B\|_2^2 / (\psi, |x|^{-1} \psi). \quad (2.12)$$

Note that the set $\mathcal{F}$ (defined in Sect. I) is not empty (see [8]).

Theorem 2.3.

$$\hat{z} = z_c. \quad (2.13)$$

Proof. Assume that $E_n \rightarrow -\infty$. We shall show that $z \geq \hat{z}$, which implies $z_c \geq \hat{z}$. By the remark above (and passing to a subsequence), we can assume that

$$\begin{aligned} \phi_n &\rightharpoonup \phi \neq 0 \text{ weakly in } H^1(\mathbb{R}^3), \\ (\phi_n, |x|^{-1} \phi_n) &\rightarrow (\phi, |x|^{-1} \phi), \\ \beta_n &\rightharpoonup \beta \text{ weakly in } L^2(\mathbb{R}^3). \end{aligned}$$

From this, $\|\beta\|_2 \leq \liminf \|\beta_n\|_2$ and $\|\phi\|_2 \leq \liminf \|\phi_n\|_2 = 1$. By (2.11)

$$0 \geq \liminf \lambda_n E_n \geq \varepsilon \|\beta\|_2^2 - z(\phi, |x|^{-1} \phi).$$

Since ϕ might not be normalized, define $\hat{\phi} = \phi / \|\phi\|_2$. Then

$$z \geq \varepsilon \|\beta\|_2^2 / (\phi, |x|^{-1} \phi) \geq \varepsilon \|\beta\|_2^2 / (\hat{\phi}, |x|^{-1} \hat{\phi}) \geq \hat{z},$$

since $(\hat{\phi}, \alpha) \in \mathcal{F}$.

On the other hand, if $z_c > \hat{z}$, then there exists $(\psi, A) \in \mathcal{F}$ such that the ratio on the right side of (2.12) is less than $\bar{z} \equiv \hat{z} + \frac{1}{2}(z_c - \hat{z})$. Define $\psi_n(x) = n^{3/2} \psi(nx)$, $A_n(x) = nA(nx)$. Then for the $\bar{z}$ just defined

$$\mathcal{E}(\psi_n, A_n) = n \{ -\bar{z}(\psi, |x|^{-1} \psi) + \varepsilon \|B\|_2^2 \},$$

which tends to $-\infty$ as $n \rightarrow \infty$. This is a contradiction since $\bar{z} < z_c$. $\square$

Remark. We have repeatedly used the facts that $z < z_c \Rightarrow E > -\infty$ and $z > z_c \Rightarrow E = -\infty$. For the case $z = z_c$ we do not know whether E is finite or $E = -\infty$. This is an open question.

Two natural questions arise. Is there a minimizing $(A, \psi) \in \mathcal{C}$ for $\mathcal{E}$ when $z < z_c$? Is there a minimizing $(A, \psi) \in \mathcal{F}$ for the ratio in (2.12) defining $z_c = \hat{z}$? The answer to both questions is yes.

Theorem 2.4. When $z < z_c$ there exists a pair $(\psi, A) \in \mathcal{C}$ such that

$$\mathcal{E}(\psi, A) = E \equiv \inf \{ \mathcal{E}(\psi', A') \mid (\psi', A') \in \mathcal{C} \}.$$

Theorem 2.5. There exists $(\psi, A) \in \mathcal{F} \equiv \{ (\psi', A') \in \mathcal{C} \mid \sigma \cdot (p - A') \psi' = 0 \}$ such that $\varepsilon \|B\|_2^2 / (\psi, |x|^{-1} \psi) = z_c = \hat{z}$ (given by (2.12)).

Proof of Theorem 2.4. Let $(\psi_n, A_n) \in \mathcal{C}$ be a minimizing sequence. By Theorem 2.2, $b_n \equiv (\psi_n, |x|^{-1} \psi_n)$ is a bounded sequence (since $z < z_c$). From (2.8) we see that $\|B_n\|_2$ and $\|\sigma \cdot (p - A_n) \psi_n\|_2$ are also bounded sequences (since $E_n < 0$). Now,

$$\|\nabla \psi_n\|_2 = \|\sigma \cdot p \psi_n\|_2 \leq \|\sigma \cdot (p - A_n) \psi_n\|_2 + \|(\sigma \cdot A_n) \psi_n\|_2.$$

However $\|(\sigma \cdot A_n)\psi_n\|_2 = \|A_n\psi_n\|_2 \leq \|A_n\|_6 \|\psi_n\|_6^{1/2} \|\psi_n\|_2^{1/2}$. By the Sobolev inequality (2.10) applied to A_n and ψ_n , (and with $\|\psi_n\|_2 = 1$),

$$\|\nabla\psi_n\|_2 \leq \|\sigma \cdot (p - A_n)\psi_n\|_2 + S^{-3/2} \|B_n\|_2 \|\nabla\psi_n\|_2^{1/2}. \quad (2.14)$$

This implies that $\|\nabla\psi_n\|_2$ is also bounded and hence that ψ_n is bounded in H^1 .

By passing to a subsequence we have

$$\begin{aligned} \psi_n &\rightharpoonup \psi \quad \text{weakly in } H^1, \\ A_n &\rightharpoonup A \quad \text{weakly in } L^6, \\ B_n &\rightharpoonup B \quad \text{weakly in } L^2, \\ \operatorname{div} A &= 0 \quad \text{and} \quad \operatorname{curl} A = B, \\ (\psi_n, |x|^{-1}\psi_n) &\rightarrow (\psi, |x|^{-1}\psi). \end{aligned} \quad (2.15)$$

The proof of the last statement is as in Theorem 2.2. Furthermore, $\psi_n A_n \rightharpoonup \psi A$ in L^2 (as in the proof of Theorem 2.1). By lower semicontinuity of the norms we obtain $E \geq \mathcal{E}(\psi, A)$. If we knew that $\|\psi\|_2 = 1$ [and hence that $(\psi, A) \in \mathcal{C}$] we would be done. However, $\|\psi\|_2 \leq 1$ by lower semicontinuity. Suppose that $\gamma = \|\psi\|_2^{-1} > 1$. Define $\hat{\psi} = \gamma\psi$. Then

$$\mathcal{E}(\hat{\psi}, A) = \gamma^2 \{ \|\sigma \cdot (p - A)\psi\|_2^2 - z(\psi, |x|^{-1}\psi) \} + \varepsilon \|B\|_2^2.$$

The term in $\{ \}$ must be negative [since $\mathcal{E}(\psi, A) \leq E < 0$]. Therefore, $E \leq \mathcal{E}(\hat{\psi}, A) < \mathcal{E}(\psi, A) \leq E$. Hence $\gamma = 1$ and the proof is complete. $\square$

Proof of Theorem 2.5. Let $(\psi_n, A_n) \in \mathcal{F}$ be a minimizing sequence. By scaling

$$\psi_n(x) \rightarrow \lambda^{3/2} \psi_n(\lambda x), \quad A_n(x) \rightarrow \lambda A_n(\lambda x),$$

we can assume that $\|B_n\|_2 = 1$. Also $\|\psi_n\|_2 = 1$ and $\|\nabla\psi_n\|_2 \leq S^{-3}$ [by (2.14) and $\sigma \cdot (p - A_n)\psi_n = 0$]. Thus ψ_n is bounded in H^1 . Again, (2.15) holds for some subsequence. By lower semicontinuity, $\|B\|_2 \leq \|B_n\|_2$ and $\gamma = \|\psi\|_2^{-1} \geq 1$. Note that $\psi \neq 0$ by the last line of (2.15). Replacing ψ by $\hat{\psi} = \gamma\psi$ we have that

$$\hat{z}/\varepsilon = \lim \{ \|B_n\|_2^2 / (\psi_n, |x|^{-1}\psi_n) \} \geq \|B\|_2^2 / (\psi, |x|^{-1}\psi) \geq \|B\|_2^2 / (\hat{\psi}, |x|^{-1}\hat{\psi}).$$

If we can show that $\sigma \cdot (p - A)\psi = 0$, we can conclude from this that $\gamma = 1$ and that (ψ, A) is a minimizing pair. However, $\nabla\psi_n \rightharpoonup \nabla\psi$ and $A_n\psi_n \rightharpoonup A\psi$ weakly in L^2 , so $0 = \sigma \cdot (p - A_n)\psi_n \rightharpoonup \sigma \cdot (p - A)\psi$ weakly in L^2 . But the weak limit of 0 can only be 0. $\square$

III. A Lower Bound for z_c

In the previous section z_c was shown to be finite (since $\mathcal{F}$ is not empty) and a formula for z_c was given. While we are unable to evaluate that formula, we shall show here that z_c is not too small. The methods of this section are completely different from those of the previous section.

Let $(\psi, A) \in \mathcal{C}$ be given and let

$$T(\psi, A) \equiv \|(p - A)\psi\|_2^2, \quad (3.1)$$

$$Q(\psi) \equiv \frac{1}{4\varepsilon} \int \{ \langle \psi, \psi \rangle(x) \}^2 dx, \quad (3.2)$$

$$T(\psi, A) \equiv \|\sigma \cdot (p - A)\psi\|_2^2 + \varepsilon \|B\|_2^2. \quad (3.3)$$

$$X(\psi, A) \equiv \frac{1}{2} T(\psi, A) / Q(\psi). \quad (3.4)$$

Lemma 3.1.

$$T(\psi, A) \geq T(\psi, A) \cdot \begin{cases} \frac{1}{2} X(\psi, A) & \text{if } X(\psi, A) \leq 1 \\ 1 - \{2X(\psi, A)\}^{-1} & \text{if } X(\psi, A) \geq 1. \end{cases} \quad (3.5)$$

Proof. Let $0 \leq t \leq 1$ and observe that

$$T(\psi, A) \geq t \|\sigma \cdot (p - A)\psi\|_2^2 + \varepsilon \|B\|_2^2.$$

Expanding the first term on the right we get

$$T(\psi, A) \geq t T(\psi, A) - t \int B(x) \cdot \langle \psi, \sigma \psi \rangle(x) dx + \varepsilon \|B\|_2^2. \quad (3.6)$$

Note that in obtaining this result we performed a partial integration in the second term which is easily justified. Minimizing the second and the third term with respect to $B(B(x) = (t/2\varepsilon) \langle \psi, \sigma \psi \rangle(x))$ we find for these two terms the lower bound

$$-t^2 Q(\psi). \quad (3.7)$$

We have used the identity

$$\langle \psi, \sigma \psi \rangle \cdot \langle \psi, \sigma \psi \rangle = \langle \psi, \psi \rangle^2, \quad \text{all } x. \quad (3.8)$$

The sum of the first term in (3.6) together with (3.7) has its maximum as a function of t at $t_0 = X(\psi, A)$. If $t_0 \leq 1$ we find

$$T(\psi, A) \geq \frac{1}{2} T(\psi, A) X(\psi, A),$$

and if $t_0 > 1$ we set $t = 1$ and get

$$T(\psi, A) \geq T(\psi, A) - Q(\psi). \quad \square$$

Lemma 3.1 provides us with two alternatives.

Alternative 1. $X(\psi, A) \geq 1$. In this case $T(\psi, A) \geq \frac{1}{2} T(\psi, A)$, and thus

$$\mathcal{E}(\psi, A) \geq \frac{1}{2} T(\psi, A) - z(\psi, |x|^{-1}\psi) \geq -\frac{1}{2} z^2. \quad (3.9)$$

Here we have used the diamagnetic inequality [4]

$$T(\psi, A) \geq T(|\psi|, 0) = \|\nabla \phi\|_2^2, \quad \text{with } \phi(x)^2 = \langle \psi, \psi \rangle(x), \quad (3.10)$$

together with the well-known hydrogenic ground state energy. We shall return to this alternative later.

Alternative 2. $X(\psi, A) < 1$. Then

$$\mathcal{E}(\psi, A) \geq \mathcal{D}(\psi, A) \equiv \frac{1}{4} T(\psi, A)^2 / Q(\psi) - z(\psi, |x|^{-1}\psi). \quad (3.11)$$

The two terms in $\mathcal{D}(\psi, A)$ scale (with x) in the same way, and hence the infimum of $\mathcal{D}$ is either 0 or $-\infty$. Let us define

$$\tilde{z}_c = \sup \{z | \inf \mathcal{D}(\psi, A) = 0\}. \quad (3.12)$$

Clearly

$$\tilde{z}_c \leq z_c. \quad (3.13)$$

Another expression for $\tilde{z}_c$ [which uses the common x -scaling of the terms in $\mathcal{D}$ and (3.10)] is

$$\tilde{z}_c = \varepsilon \inf_{\phi} \|\phi\|_2^2 \|\nabla \phi\|_2^4 \|\phi\|_4^{-4} (\phi, |x|^{-1} \phi)^{-1}. \quad (3.14)$$

Here ϕ is an ordinary real valued function in $H^1(\mathbb{R}^3)$.

As an aside, it is worth mentioning that the fact that $z_c \geq \tilde{z}_c$ [given by (3.14)] – but not Lemma 3.1 – can be derived directly from the formula (2.12). If $\sigma \cdot (p - A)\psi = 0$, then $0 = \int |\sigma \cdot (p - A)\psi|^2 = \int |(p - A)\psi|^2 - \int B \cdot \langle \psi, \sigma \psi \rangle$. (A justified integration by parts was used in the last term.) Using the Schwarz inequality on the last term, and (3.8), we have

$$T(\psi, A) \leq \|B\|_2 Q(\psi)^{1/2} (4\varepsilon)^{1/2}. \quad (3.15)$$

Equation (3.14) then follows from (3.15), (3.10) and formula (2.12).

Our next goal is to find a lower bound to the right side of (3.14), which we shall call z_c^L :

$$z_c^L \leq \tilde{z}_c \leq z_c. \quad (3.16)$$

(Of course one can try to compute the infimum in (3.14) directly – which leads to an interesting differential equation.) First note that

$$\|\nabla \phi\|_2 \|\phi\|_2 \geq (\phi, |x|^{-1} \phi), \quad (3.17)$$

which is the uncertainty principle and follows from the hydrogen ground state by scaling in x . Hence

$$\tilde{z}_c \geq z_c^L \equiv \varepsilon \inf_{\phi} \|\nabla \phi\|_2^3 \|\phi\|_4^{-4} \|\phi\|_2. \quad (3.18)$$

The minimization problem in (3.18) is equivalent to the following. Let $e < 0$ be the ground state energy of $-\Delta - V(x)$. In [7] it is proved that

$$|e|^{1/2} \leq L_{\frac{3}{2}, 3}^1 \|V\|_2^2. \quad (3.19)$$

$L_{\frac{3}{2}, 3}^1$ is obtained by solving an ordinary differential equation [7] and is found numerically (to 3 significant figures) to be

$$L_{\frac{3}{2}, 3}^1 = 0.0135. \quad (3.20)$$

By choosing $V(x) = C\phi(x)^2$ one deduces from (3.19), that $\|\nabla \phi\|_2^2 \geq C\|\phi\|_4^4 - C^4(L_{\frac{3}{2}, 3}^1)^2\|\phi\|_4^8$ when $\|\phi\|_2 = 1$. Optimizing with respect to C and inserting the result in (3.18) gives

$$z_c^L = \varepsilon(3/4)^{3/2} \{2L_{\frac{3}{2}, 3}^1\}^{-1} \geq (24.0)\varepsilon. \quad (3.21)$$

This lower bound to $\tilde{z}_c$ is surprisingly accurate. Inserting the function $\phi(x) = \exp(-|x|)$ in (3.14) yields

$$z_c^L \leq \tilde{z}_c \leq 8\pi\varepsilon = (25.1)\varepsilon. \quad (3.22)$$

Recalling that $\varepsilon = \{8\pi\alpha^2\}^{-1} = 747.2$, we have that

$$z_c \geq z_c^L \geq 17,900. \quad (3.23)$$

In [8] a solution to $\sigma \cdot (p - A)\psi = 0$ is found which, when inserted into (2.12), yields

$$z_c \leq 9\pi^3\varepsilon = (279)\varepsilon = 208,000. \quad (3.24)$$

So far we have shown that if $z \leq \tilde{z}_c$ then alternative 2 above is irrelevant, for otherwise we should conclude that $E \geq 0$, which is false. Our next goal is to find a lower bound for E , and we shall do so under the slightly stronger condition that $z \leq z_c^L$. To this end, we need only consider $(\psi, A) \in \mathcal{C}$ such that $X(\psi, A) \geq 1$. By Lemma 3.1, a lower bound, $\tilde{E}$, for E is given by

$$E \geq \tilde{E} = \inf \{T(\psi, A) - Q(\psi) - z(\psi, |x|^{-1}\psi)\} \quad (3.25)$$

under the conditions $(\psi, A) \in \mathcal{C}$ and $T(\psi, A) \geq 2Q(\psi)$.

The problem posed by (3.25) is too difficult (in particular it is not clear that $A=0$ is an optimal choice). Therefore we seek a lower bound to the right side of (3.25) as follows. Recall that $T(\psi, A)$ satisfies (for $\|\psi\|_2 = 1$)

$$T(\psi, A) \geq [4z_c^L Q(\psi)]^{2/3}, \quad (3.26)$$

which follows from (3.18) and (3.10),

$$T(\psi, A) \geq (\psi, |x|^{-1}\psi)^2, \quad (3.27)$$

$$T(\psi, A) \geq 2Q(\psi). \quad (3.28)$$

Define $\tau(\psi)$ by

$$\tau(\psi) = \max \{\text{right sides of (3.26), (3.27), (3.28)}\}.$$

Then, if we define E^L by

$$E^L = \inf_{\psi} \{\tau(\psi) - Q(\psi) - z(\psi, |x|^{-1}\psi)\} \quad (3.29)$$

(with $\|\psi\|_2 = 1$), we have that

$$E^L \leq \tilde{E} \leq E. \quad (3.30)$$

The problem posed by (3.29) is, in fact, algebraic. It is solved in Appendix B with the result that for all $z \leq z_c^L$

$$E^L = -\frac{1}{4}\left(\frac{4}{3}\right)^3 (z_c^L)^2 [3\gamma - 2 + 2(1-\gamma)^{3/2}], \quad (3.31)$$

where $\gamma = \frac{3}{4}z/z_c^L$.

When the right side of (3.31) is Taylor expanded for small z , the leading two terms are

$$\approx -\frac{1}{4}z^2 - \frac{1}{32z_c^L}z^3. \quad (3.32)$$

On the other hand using Taylor's theorem with remainder and taking the maximum of $d^3 E^L/dz^3$ in the interval $[0, z]$, we can derive a lower bound to (3.31) for all $z \leq z_c^L$, which agrees with (3.32) to the first two orders:

$$E \geq -\frac{1}{4}z^2 - (32z_c^L)^{-1}z^3(1-\gamma)^{-3/2}. \quad (3.33)$$

A crude upper bound for E can be obtained with the trial function

$$\begin{aligned} \psi(r) &= \begin{pmatrix} \phi(r) \\ 0 \end{pmatrix}, \quad \phi(r) = (z/2)^{3/2} \pi^{-1/2} e^{-zr/2}, \\ A &= \frac{1}{2} \alpha^2 z^3 (2/3)^4 e^{-zr/2} (-y, x, 0). \end{aligned} \quad (3.34)$$

(This choice does not satisfy $\operatorname{div} A = 0$, but that does not matter.) A computation with this (ψ, A) gives

$$E \leq -\frac{1}{4}z^2 - \frac{1}{2}\left(\frac{2}{3}\right)^8 z^3 \alpha^2 + 2^4 3^{-8} z^4 \alpha^4. \quad (3.35)$$

Remark. We do not know whether E diverges as $z \rightarrow z_c$. Of course, E is an upper semicontinuous, monotone decreasing function of z , so $E(z_c) = \lim_{z \rightarrow z_c} E(z)$.

IV. A Single Electron Atom in a Magnetic Field of Constant Direction

In the previous two sections we considered a single electron atom in an arbitrary magnetic field and showed that z_c is finite (but huge) and estimated the shift in the ground state energy for $z < z_c^L$. The magnetic field that causes the energy to diverge when $z > z_c$ has to be highly contorted (which is consistent with the example given in [8]). If, on the other hand, certain constraints are placed on B near the nucleus, the divergence will not occur and z_c will be infinite.

In this section we display one such condition—namely that B has a constant direction (but not necessarily constant magnitude) near the nucleus. This is one possible version of the external field problem and is relevant for astrophysics. We shall content ourselves with showing merely that z_c is infinite and will not bother to try to find a good estimate on the energy; in fact we shall obtain $E \geq -(\text{const})(1+z^4)$ for all z . The crucial point, of course, is that the bound is independent of B (but it does depend on the size of the region in which the direction of B is constant).

It should be noted that something a bit stronger is actually proved in the following. Namely for *any* B the energy will be bounded below if we replace the troublesome term $\sigma \cdot B$ by $\sigma_3 B_3$. Such a replacement is physically meaningful only when $B_1 = B_2 = 0$.

Let R be a fixed radius and assume that inside the ball K_R of radius R centered at the origin (which is also the location of the nucleus)

$$B(x) = (0, 0, b(x)). \quad (4.1)$$

(The choice of the 3 direction is arbitrary.) $b(x)$ can be anything inside K_R and $B(x)$ can also be anything outside of K_R .

Let $\psi(x)$ be given on $\mathbb{R}^3$, and we want to localize it inside and outside K_R . Define η_1, η_2 both C^∞ such that $\eta_1(x) = 1$ for $x \in K_{R/2}$ and $\eta_1(x) = 0$ for $x \notin K_R$ and

$\eta_1(x)^2 + \eta_2(x)^2 = 1$. Also define $\psi_i(x) = \eta_i(x)\psi(x)$, $i = 1, 2$. Thus $\psi^2 = \psi_1^2 + \psi_2^2$. It is easy to see that

$$\|\sigma \cdot (p - A)\psi_1\|_2^2 + \|\sigma \cdot (p - A)\psi_2\|_2^2 = (\psi, f\psi) + \|\sigma \cdot (p - A)\psi\|_2^2, \quad (4.2)$$

where $f = (\nabla\eta_1)^2 + (\nabla\eta_2)^2$. (The cross terms cancel.) We can easily choose η_i such that $f(x) \leq dR^{-2}$ for some constant, d . Hence we get for $(\psi, A) \in \mathcal{C}$,

$$\mathcal{E}'(\psi, A) \geq \|\sigma \cdot (p - A)\psi_1\|_2^2 - z(\psi_1, |x|^{-1}\psi_1) + \varepsilon\|B\|_2^2 - d/R^2 - 2z/R. \quad (4.3)$$

Here we used the fact that $(\psi_2, |x|^{-1}\psi_2) \leq 2/R$, since $\psi_2(x) = 0$ for $|x| < R/2$ and $\|\psi_2\|_2 \leq 1$.

From now on we drop the subscript 1 and denote ψ_1 by ψ (with $\|\psi\|_2 \leq 1$). Define

$$T_3(\psi, A) = \|(p_3 - A_3)\psi\|_2^2 \geq \|p_3|\psi|\|_2^2 \equiv T_3(\psi), \quad (4.4)$$

$$T_\perp(\psi, A) = \|(p_\perp - A_\perp)\psi\|_2^2 \geq \|p_\perp|\psi|\|_2^2 \equiv T_\perp(\psi), \quad (4.5)$$

where $p_\perp = (p_1, p_2)$, etc. [The inequality (3.10) holds in any dimensions.] Since $B(x)$ is given by (4.1) on the set where $\varphi(x) \neq 0$, we have

$$\|\sigma \cdot (p - A)\psi\|_2^2 = T_3(\psi, A) + T_\perp(\psi, A) - \int b\langle\psi, \sigma_3\psi\rangle, \quad (4.6)$$

which can be rewritten as

$$T_3(\psi, A) + U_\perp(\psi, A), \quad (4.7)$$

where

$$U_\perp(\psi, A) \equiv \|\sigma_\perp \cdot (p_\perp - A_\perp)\psi\|_2^2. \quad (4.8)$$

Consider E' , which is defined to be the infimum over $(\psi, A) \in \mathcal{C}$ of

$$\mathcal{E}'(\psi, A) = T_3(\psi) + U_\perp(\psi, A) + \varepsilon\|b\|_2^2 - z(\psi, |x|_R^{-1}\psi), \quad (4.9)$$

where $|x|_R^{-1} = |x|^{-1}$ if $|x| \leq R$ and zero otherwise. Here b is defined to be the 3-component of $B = \text{curl} A$, even if B does not point in the 3-direction (note that $\|B\|_2 \geq \|b\|_2$, and that (4.6)–(4.8) is still true, namely $U_\perp = T_\perp - \int b\langle\psi, \sigma_3\psi\rangle$). It is obvious that

$$E \geq E' - d/R^2 - 2z/R. \quad (4.10)$$

To analyze E' we observe that each of the four terms in (4.9) involves a 3-dimensional integral, and $\int d^3x = \int dx_3 \int dx_\perp$. Think of $\psi, A, B, |x|_R^{-1}$ as functions of $x_\perp$ parameterized by x_3 . Then

$$\mathcal{E}'(\psi, A) = T_3(\psi) + \int dx_3 \mathcal{E}''(\psi, A), \quad (4.11)$$

$$\mathcal{E}''(\psi, A) = \int dx_\perp |\sigma_\perp \cdot (p_\perp - A_\perp)\psi|^2 + \varepsilon \int dx_\perp b^2 - z \int_{D(x_3)} dx_\perp \langle\psi, \psi\rangle (x_\perp^2 + x_3^2)^{-1/2}, \quad (4.12)$$

where $D(x_3)$ is the domain in $x_\perp$ given by

$$x_\perp^2 \leq R^2 - x_3^2. \quad (4.13)$$

To analyze (4.12) we utilize the t trick of Sect. III. For each value of x_3 , let $t(x_3)$ be chosen to satisfy $0 \leq t(x_3) \leq 1$. Replace $|\sigma \cdot (p_\perp - A_\perp)\psi|^2$ in (4.12) by $t(x_3)$ times

this quantity and use (3.10) to obtain the lower bound on this first term:

$$t(x_3)T_{\perp}(\psi) - t(x_3) \int |b| \langle \psi, \psi \rangle dx_{\perp}.$$

[Here $T_{\perp}$ means $\int dx_{\perp} (\nabla_{\perp} |\psi|)^2$.] We used $|\langle \psi, \sigma \psi \rangle| = \langle \psi, \psi \rangle$.

Now minimize with respect to b and then maximize with respect to $t(x_3)$, as in Sect. III. For the first two terms on the right side of (4.12) we obtain the bound

$$\min \{ (J_{\perp})^2 / 4\epsilon, \epsilon (T_{\perp} / J_{\perp})^2 \} \quad (4.14)$$

with $(J_{\perp})^2 = \int dx_{\perp} \langle \psi, \psi \rangle^2$.

The last term in (4.12) can be bounded below as $-zJ_{\perp}W(x_3)$, and

$$W(x_3)^2 = \int_{D(x_3)} (x_{\perp}^2 + x_3^2)^{-1} dx_{\perp} = \begin{cases} 2\pi \ln(R/|x_3|) & \text{for } |x_3| < R \\ 0 & \text{for } |x_3| \geq R. \end{cases} \quad (4.15)$$

To bound (4.14) below, the Sobolev inequality in $\mathbb{R}^2$ is used:

$$T_{\perp} \geq S(J_{\perp})^2 / g(x_3)^2, \quad (4.16)$$

where

$$g(x_3)^2 = \int dx_{\perp} \langle \psi, \psi \rangle. \quad (4.17)$$

(The constant S can be found in [7].)

Substituting (4.14)–(4.17) in (4.12),

$$\mathcal{E}''(\psi, A) \geq J_{\perp}^2 \min \{ (4\epsilon)^{-1}, S^2 \epsilon g(x_3)^{-4} \} - zJ_{\perp}W(x_3). \quad (4.18)$$

Since $J_{\perp} = J_{\perp}(x_3)$ is unknown, we simply minimize (4.18) with respect to $J_{\perp}$ and obtain

$$\mathcal{E}''(\psi, A) \geq -\frac{1}{4}z^2W(x_3)^2 \cdot \max \{ 4\epsilon, g(x_3)^4 / S^2 \epsilon \}. \quad (4.19)$$

According to (4.11), (4.19) must be integrated over x_3 . Since we do not know which term in the $\max \{ \cdot \}$ in (4.19) holds for any given x_3 , we shall simply take the sum of the two. The first yields

$$-\epsilon z^2 \int_{-R}^R dx_3 W(x_3)^2 = -4\pi \epsilon z^2 R. \quad (4.20)$$

To control the second possibility we invoke the $T_3(\psi)$ term in (4.11). An application of the Schwarz inequality [12] gives

$$T_3(\psi) \geq \int dx_3 (dg(x_3)/dx_3)^2 = \|g'\|_2^2. \quad (4.21)$$

It is also a fact that for all x_3 and $g \in L^2(\mathbb{R}^1)$

$$g(x_3)^4 \leq \|g\|_2^2 \|g'\|_2^2 \leq T_3(\psi). \quad (4.22)$$

[This follows from $g(x)^2 = 2 \int_{-\infty}^x gg'$ and $g(x)^2 = -2 \int_x^{\infty} gg'$. Hence $g(x)^2 \leq \int_{-\infty}^{\infty} |gg'|$. We recall also that $\|g\|_2^2 = \int \langle \psi, \psi \rangle d^3x = 1$.] Inserting (4.20)–(4.22) in (4.11), we obtain the following lower bound for the second possibility in (4.19).

$$T_3(\psi) \{ 1 - \frac{1}{4}z^2 S^{-2} \epsilon^{-1} \int W^2 \} = T_3(\psi) \{ 1 - \pi R z^2 / S^2 \epsilon \}. \quad (4.23)$$

While the value of $T_3(\psi)$ is unknown, the term (4.23) can be eliminated by the following trick. Call R_0 the original radius inside of which (4.1) holds. A fortiori,

(4.1) holds for any $R < R_0$. If $\{\bar{\cdot}\}$ in (4.23) is nonnegative, use R_0 . Otherwise, use

$$R = S^2 \varepsilon / \pi z^2.$$

Then (4.23) ≥ 0 and can be ignored.

Combining all the terms we obtain

$$E \geq -d/R^2 - 2z/R - 4\pi \varepsilon z^2 R \quad (4.24)$$

with

$$R = \min \{R_0, S^2 \varepsilon / \pi z^2\}. \quad (4.25)$$

Appendix A

In the following, $\mathscr{D}'$ is the space of distributions.

Theorem A.1. *Let $B \in L^2(\mathbb{R}^3)$ be a given vector field and let $\operatorname{div} B = 0$ in $\mathscr{D}'$. Define the vector field*

$$A(x) = \frac{1}{4\pi} \int |x-y|^{-3} (x-y) \times B(y) dy. \quad (A.1)$$

Then:

- (a) $A \in L^6(\mathbb{R}^3)$ and $\operatorname{curl} A = B$, $\operatorname{div} A = 0$ in $\mathscr{D}'$.
- (b) The distribution $\partial_i A_j$ is an L^2 -function and we have the formula

$$\sum_i \int |\nabla A_i|^2 d^3x = \int B^2 dx.$$

- (c) The $A(x)$ given by (A.1) is the only vector field having the three properties in (a) above.

Proof. Let us write $A = T(B)$. The kernel in (A.1) is bounded by $|x-y|^{-2}$ and $|x|^{-2} \in L_w^{3/2}$. Let V be a vector field in $L^p(\mathbb{R}^3)$ with $1 < p < 3$. By the weak Young inequality, $T(V) \in L^r$ where $1/3 + 1/r = 1/p$, and $\|T(V)\|_r \leq C_p \|V\|_p$ for a suitable constant C_p . By Fubini's theorem

$$(W, T(V)) = (T(W), V) \quad (A.2)$$

when $W \in L^q$ and $V \in L^p$, with $q = r'$, $1/q = 4/3 - 1/p$. In (A.2), (W, U) means $\sum_{i=1}^3 \int \overline{W_i(x)} U_i(x) dx$.

Now we apply (A.2) to $V = B$ and $W = \nabla f$ with $f \in C_0^\infty(\mathbb{R}^3)$,

$$T(W) = -\frac{1}{4\pi} \operatorname{curl} \{ |x|^{-1} * \nabla f \} = -\frac{1}{4\pi} \operatorname{curl} \operatorname{grad} \{ |x|^{-1} * f \} = 0.$$

(The first equality, namely exchanging integration and differentiation, follows by dominated convergence.) Then $(\nabla f, A) = 0$ for all $f \in C_0^\infty$, and hence $\operatorname{div} A = 0$ in $\mathscr{D}'$. A second application of (A.2) is to $V = B$ and $W = \operatorname{curl} G$, with $G \in C_0^\infty$.

$$T(W) = -\frac{1}{4\pi} \operatorname{curl} \operatorname{curl} \{ |x|^{-1} * G \} = -G - \nabla \operatorname{div} \frac{1}{4\pi} \{ |x|^{-1} * G \} \equiv -G - \nabla g.$$

Then $(W, A) = (-G - \nabla g, B) = (-G, B) - (\nabla g, B)$. We claim that $(\nabla g, B) = 0$, which will imply that $\operatorname{curl} A = B$ in $\mathscr{D}'$. While $g \in C^\infty$ it does not generally have compact support; otherwise we would have $(\nabla g, B) = 0$ since $\operatorname{div} B = 0$ in $\mathscr{D}'$.

However $B \in L^2$, and therefore $(\nabla g, B) = \lim_{R \rightarrow 0} (\nabla(gf_R), B)$, where $f_R(x) = f(Rx)$ and $f(x)$ is a C_0^∞ function satisfying $f(x) = 1$ for $|x| < 1$, $f(x) = 0$ for $|x| > 2$. Since $(\nabla(gf_R), B) = 0$, we have the desired result, and (a) is proved.

To prove (b) we define $T_j(B) = i\partial_j T(B) = iT(\partial_j B)$ for B smooth and of compact support. It is a standard result about the Riesz transform that $T_j(B)$ has a bounded extension to $L^2(\mathbb{R}^3)$, see [10], so we can assume merely that $B \in L^2$. Furthermore T_j is selfadjoint. Now for any vector field V in $L^2(\mathbb{R}^3)$ we have

$$\sum_j (T_j(B), T_j(V)) = \sum_j (B, T_j^2(V)) = (B, V). \quad (\text{A.3})$$

Indeed, when V is smooth and of compact support $\sum_j T_j^2(V)(x) = V(x) + \frac{1}{4\pi} \nabla \operatorname{div} \{ |x|^{-1} * V \}(x)$. Using the previous approximation argument (namely $g \rightarrow gf_R$) and the fact that $\operatorname{div} B = 0$ in $\mathcal{D}'$ gives (A.3). Since T_j is bounded, (A.3) is true for all $V \in L^2(\mathbb{R}^3)$. Hence, by setting $V = B$, (b) is also proven.

To prove (c), suppose there were another $\tilde{A}$ with the properties in (a) and let $\alpha = \tilde{A} - A$. Then $\alpha \in L^6$, $\operatorname{curl} \alpha = 0$, $\operatorname{div} \alpha = 0$ in $\mathcal{D}'$. Let $j_\varepsilon(x)$ be a C_0^∞ approximation to the identity and $\alpha_\varepsilon = j_\varepsilon * \alpha$. It is easy to see that $\alpha_\varepsilon \in C^\infty$, $\operatorname{div} \alpha_\varepsilon = 0$, $\operatorname{curl} \alpha_\varepsilon = 0$ and $\alpha_\varepsilon(x) \rightarrow 0$ as $|x| \rightarrow \infty$. From this, $\Delta \alpha_\varepsilon = -\operatorname{curl} \operatorname{curl} \alpha_\varepsilon + \operatorname{grad} \operatorname{div} \alpha_\varepsilon = 0$. So each component of α_ε is harmonic, but since $\alpha_\varepsilon \rightarrow 0$ at ∞ , α_ε must be zero for all $\varepsilon > 0$. But as $\varepsilon \rightarrow 0$, $\alpha_\varepsilon \rightarrow \alpha$ (in L^6 and in $\mathcal{D}'$), so $\alpha = 0$. $\square$

Theorem A.2. For any $A \in L^6(\mathbb{R}^3)$ and $\psi \in L^2(\mathbb{R}^3)$, $\|\sigma \cdot (p - A)\psi\|_2 < \infty$ implies $\psi \in H^1(\mathbb{R}^3)$.

Proof. Observe that by assumption

$$\sigma \cdot p\psi = \sigma \cdot A\psi + u, \quad u \in L^2, \quad \text{and} \quad A\psi \in L^{3/2}$$

by Hölder's inequality ($A \in L^6$, $\psi \in L^2$). Since $(\sigma \cdot p)^{-1} = \sigma \cdot p|p|^{-2}$, we find

$$\psi = \frac{i}{4\pi} \int |x - y|^{-3} \sigma \cdot (x - y) [(\sigma \cdot A\psi)(y) + u(y)] dy.$$

Again, by the weak Young inequality, $\psi = v_1 + v_2$, $v_1 \in L^3$, $v_2 \in L^6$ which implies (since $\psi \in L^2$) $\psi \in L^2 \cap L^3$. Hence $A\psi \in L^2$ (again by Hölder's inequality) and thus $\sigma \cdot p\psi \in L^2$.

Appendix B: Proof of Eq. (3.31)

Given ψ , define

$$S(\psi) = (\psi, |x|^{-1}\psi).$$

As ψ ranges over all functions satisfying $\|\psi\|_2 = 1$, $Q(\psi)$ and $S(\psi)$ independently take on all values between 0 and ∞ . Therefore we are entitled to think of S and Q simply as an unknown pair of positive numbers.

According to (3.29), then, we have to minimize $e = \tau - Q - zS$ under the conditions $\tau \geq 2Q$, $\tau \geq S^2$, $\tau \geq KQ^{2/3}$ with $K^{3/2} = 4z_c^L$. There are two cases:

case (a): $2Q^{1/3} \geq K$ or $Q \geq 2(z_c^L)^2$,

case (b): $2Q^{1/3} < K$.

If case (a) holds, we set $\tau = 2Q$, $S^2 = 2Q$, and then $\tau - Q - zS \geq 0$, since $z \leq z_c^L$ and $Q \geq 2(z_c^L)^2$. If case (b) holds then, similarly,

$$E^L = \min \{ KQ^{2/3} - Q - zK^{1/2}Q^{1/3} | Q \leq 2(z_c^L)^2 \}.$$

Change the variable to $Q \equiv 2(z_c^L)^2 x^3$. Then E^L is the minimum of

$$-2(z_c^L)^2 [x^3 - 2x^2 + \frac{4}{3}\gamma x]$$

subject to $0 \leq x \leq 1$ and $\gamma = \frac{3}{4}z/z_c^L \leq \frac{3}{4}$. The minimum occurs at

$$x = \frac{2}{3}[1 - (1 - \gamma)^{1/2}]$$

and yields (3.31) for the lower bound.

Acknowledgement. It is a pleasure to acknowledge valuable discussions at the Aspen Center for Physics with David Boulware and Lowell Brown.

References

1. Avron, J., Herbst, I., Simon, B.: Schrödinger operators with magnetic fields: III. Atoms in homogeneous magnetic field. Commun. Math. Phys. **79**, 529–572 (1981)
2. Remark 3 in Brezis, H., Lieb, E.H.: Minimum action solution of some vector field equations. Commun. Math. Phys. **96**, 97–113 (1984)
3. Daubechies, I., Lieb, E.H.: One-electron relativistic molecules with Coulomb interaction. Commun. Math. Phys. **90**, 497–510 (1983)
4. Kato, T.: Schrödinger operators with singular potentials. Israel J. Math. **13**, 135–148 (1972)
5. Lieb, E.H.: On the lowest eigenvalue of the Laplacian for the intersection of two domains. Invent. Math. **74**, 441–448 (1983)
6. Lieb, E.H., Loss, M.: Stability of Coulomb systems with magnetic fields: II. The many-electron atom and the one-electron molecule. Commun. Math. Phys. **104**, 271–282 (1986)
7. Lieb, E.H., Thirring, W.: Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In: Studies in mathematical physics, essays in honor of Valentine Bargmann. Lieb, E.H., Simon, B., Wightman, A.S. (eds.). Princeton, NJ: Princeton University Press 1976
8. Loss, M., Yau, H.T.: Stability of Coulomb systems with magnetic fields: III. Zero energy bound states of the Pauli operator. Commun. Math. Phys. **104**, 283–290 (1986)
9. Michel, F.C.: Theory of pulsar magnetospheres. Rev. Mod. Phys. **54**, 1–66 (1982)
10. Stein, E.M.: Singular integrals and differentiability properties of functions. Princeton, NJ: Princeton University Press 1970
11. Straumann, N.: General relativity and relativistic astrophysics. Berlin, Heidelberg, New York, Tokyo: Springer 1984
12. Hoffmann-Ostenhof, M., Hoffmann-Ostenhof, T.: Schrödinger inequalities and asymptotic behavior of the electron density of atoms and molecules. Phys. Rev. A **16**, 1782–1785 (1977)
13. Avron, J., Herbst, I., Simon, B.: Schrödinger operators with magnetic fields: I. General Interactions. Duke Math. J. **45**, 847–883 (1978)

Communicated by A. Jaffe

Received October 2, 1985; in revised form January 2, 1986

Stability of Coulomb Systems with Magnetic Fields

II. The Many-Electron Atom and the One-Electron Molecule

Elliott H. Lieb* and Michael Loss**

Departments of Mathematics and Physics, Princeton University, Jadwin Hall, P.O. Box 708, Princeton, NJ 08544, USA

Abstract. The analysis of the ground state energy of Coulomb systems interacting with magnetic fields, begun in Part I, is extended here to two cases. Case A: The many electron atom; Case B: One electron with arbitrarily many nuclei. As in Part I we prove that stability occurs if $z\alpha^{12/7} < \text{const}$ (in case A) and $z\alpha^2 < \text{const}$ (in case B), ($z|e|$ = nuclear charge, α = fine structure constant), but a new feature enters in case B. There one *also* requires $\alpha < \text{const}$, regardless of the value of z .

I. Introduction

In the first paper in this series [1] the question of the stability of atoms and molecules in the presence of magnetic fields was raised, and it was answered in the case of the one-electron atom of arbitrary nuclear charge $z|e|$. In the present paper the stability question will be answered in two other cases:

- (A) The many electron atom,
- (B) The one-electron molecule.

Unfortunately, the stability of the many-electron, many-nucleus system is still an open question.

The reader is referred to the introduction in [1] for the motivation and physical interpretation of this problem. The mathematical essence of the problem is that we want to decide whether or not the energy functional

$$\begin{aligned} \mathcal{E}(\psi, A, \underline{R}, \underline{z}) \equiv & \sum_{j=1}^N \int |\sigma_j \cdot (p_j - A(x_j))\psi|^2 dx + \varepsilon \int B(x)^2 dx \\ & + (\psi, V(X, \underline{R}, \underline{z})\psi) \end{aligned} \quad (1.1)$$

* Work partially supported by U.S. National Science Foundation grant PHY-8116101-A03

** Work partially supported by U.S. and Swiss National Science Foundation Cooperative Science Program INT-8503858.

Current address: Institut für Mathematik, FU Berlin, Arnimallee 3, D-1000 Berlin 33

is bounded below by a suitable constant. The three terms in (1.1) are the electronic kinetic energy, the magnetic field energy and the Coulomb energies respectively. The notation is the following:

The energy unit is 4 Rydbergs $= 2mc^2\alpha^2$ and $1/\varepsilon = 8\pi\alpha^2$, with $\alpha = e^2/\hbar c = 1/137$ being the fine structure constant. The charge unit is $|e|$.

$\psi = \psi(x_1, \dots, x_N, s_1, \dots, s_N)$ is an arbitrary N particle, antisymmetric (electron) wave function. The particle spatial and spin coordinates are x, s with $s = \pm 1$. X denotes the collection $(x_1, \dots, x_N)$. The σ_j^i , $j = 1, 2, 3$ denote the Pauli spin matrices. ψ is assumed to be normalized

$$1 = \|\psi\|_2^2 = (\psi, \psi) = \sum_{s_1 \dots s_N} \int d^{3N}X |\psi(X, s_1, \dots, s_N)|^2. \quad (1.2)$$

$A(x)$ is a vector potential and $B = \text{curl } A$ is the magnetic field which is assumed to be in $L^2(\mathbb{R}^3)$. As explained in [1], for any $B \in L^2$, A exists and is uniquely specified by

$$\text{curl } A = B, \text{div } A = 0, A \in L^6(\mathbb{R}^3). \quad (1.3)$$

The first term in (1.1) is the electron kinetic energy. For particle j it is

$$\|\sigma_j \cdot (p_j - A)\psi\|_2^2 = \|(p_j - A)\psi\|_2^2 - (\psi, \sigma_j \cdot B\psi). \quad (1.4)$$

The Coulomb term is

$$\begin{aligned} V(X, R, z) = & \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} + \sum_{1 \leq i < j \leq K} z^i z^j |R_i - R_j|^{-1} \\ & - \sum_{i=1}^N \sum_{j=1}^K z^j |x_i - R_j|^{-1}. \end{aligned} \quad (1.5)$$

Here we assume that there are K fixed nuclei of charges $z^j|e|$ and distinct locations $R_j \in \mathbb{R}^3$, $j = 1, \dots, K$. The z 's and R 's will be denoted collectively by z and R . The first term in (1.5) is the electronic repulsion, the second is the nuclear repulsion and the third is the electron-nuclear attraction.

It is useful to have the following notation

$$\tau(\psi, A) \equiv \sum_{j=1}^N \|\sigma_j \cdot (p_j - A)\psi\|_2^2 + \varepsilon \|B\|_2^2, \quad (1.6)$$

$$T(\psi, A) \equiv \sum_{j=1}^N \|(p_j - A)\psi\|_2^2, \quad (1.7)$$

$$W(\psi, R, z) \equiv -(\psi, V(X, R, z)\psi). \quad (1.8)$$

We assume $B \in L^2$ and that (1.2) is satisfied. Then, as proved in [1] (with a slight modification to handle the N -coordinate case), in order to make sense of τ and W it is necessary and sufficient to have $\psi \in H^1(\mathbb{R}^{3N})$, i.e. ψ and all its first derivatives are in L^2 . The class of all pairs (ψ, A) satisfying the above [and also with ψ normalized as in (1.2)] is denoted by $\mathcal{C}$.

The energy of our system is defined to be

$$E \equiv \inf \{ \mathcal{E}(\psi, A, R, z) | (\psi, A) \in \mathcal{C}, \text{ all } R \}. \quad (1.9)$$

This infimum includes an infimum over R .

From [1] we know that if any single z^j satisfies $z^j > z_c$ (which is evaluated in [1] and which is proportional to α^{-2}), then $E = -\infty$, simply by moving $N-1$ electrons and the other $K-1$ nuclei to infinity. Therefore z_c for the full problem (1.1) is finite. (When $K > 1$, z_c is defined to be the largest z such that E is finite whenever all the $z^j < z$.) Our goal here is to show that z_c is not too small for (1.1). Three cases have to be distinguished.

(A) One nucleus (with $R_1 = 0$ and $z^1 \equiv z$) and an arbitrary number, N , of electrons. In Sect. II we find some $\tilde{z}_c$, which is *independent of* N , such that E is finite when $z < \tilde{z}_c$. We also find some $z_c^L < \tilde{z}_c$ for which we can give a lower bound to E (called E^L) when $z < z_c^L$. Both z_c^L and E^L are *independent of* N . The bound on z_c is

$$z_c > z_c^L \geq -\frac{1}{4} + (0.158)\alpha^{-12/7}. \quad (1.10)$$

Note the exponent 12/7. Is it possible that this can be replaced by 2, as in the one-electron case? We do not know. While our bound on $\tilde{z}_c$ utilizes the electronic Coulomb repulsion in (1.5), we conjecture that the repulsion is not really necessary. This is an interesting open problem.

(B) One electron and an arbitrary number, K , of nuclei. In Sect. III we find, as in case (A), $z_c^L < \tilde{z}_c < z_c$ (with z_c^L and $\tilde{z}_c$ independent of K and proportional to α^{-2}) such that E is finite for $z < z_c$. We also derive a lower bound $E^L < E$ when $z < z_c^L$. However, an important new feature enters here: These results also require that

$$\alpha < \alpha_c \quad (1.11)$$

for some α_c (which is shown to satisfy $0.32 < \alpha_c < 6.7$). In other words, two conditions are required for stability,

$$z^j \alpha^2 \text{ small (all } j) \text{ and } \alpha \text{ small.} \quad (1.12)$$

This situation is reminiscent of the relativistic stability problem [2–4], except that there the requirement is $z^j \alpha$ small and α small. It is interesting to note that there are other indications [5, 6] that the stability of field theory requires a bound on the coupling constant (apart from a bound on z). We shall also prove that the requirement (1.11) for stability is real; it is not an artifact of our proof.

(C) Many electrons and many nuclei. We are unable to solve this problem, but the goal would be to prove that E is finite provided $z^j \alpha^2$ is small (all j) and α is small, and that E is then bounded below by $-(\text{const})(N + K)$.

II. Basic Strategy

The following sections are full of technical details, but the common strategy (similar to that used in [1]) is simple. Let us outline it here. Note that the following steps can be carried out even for the full problem, (C), to give an N and K dependent bound on z_c . It is only in cases A and B that we can eliminate this dependence.

The quantities $\tau(\psi, A)$ and $T(\psi, A)$ were defined in (1.6), (1.7); the following quantity Q is also needed. Let $\varrho(x)$ be the one-particle density associated with ψ :

$$\varrho_\psi(x) = \sum_{j=1}^N \sum_{s_1, \dots, s_N} \int |\psi(X, s_1, \dots, s_N)|^2 d^{3N-3} X^j. \quad (2.1)$$

(X^j means all N variables *except* x_j .) Of course, for fermions we do not have to sum on j . Merely take $j = 1$ and then multiply by N . The general expression (2.1) is used because much of the following holds for any statistics (i.e. without symmetry). Then define Q by

$$Q(\psi) = (1/4\varepsilon) \int \varrho_\psi(x)^2 dx = (1/4\varepsilon) \|\varrho_\psi\|_2^2. \quad (2.2)$$

Another important quantity is the quantum ground state energy when the $\sigma \cdot B$ and the $\varepsilon \int B^2$ terms are eliminated:

$$E^q(\underline{z}) = \inf \{ T(\psi, A) - W(\psi, R, \underline{z}) \mid (\psi, A) \in \mathcal{C}, \text{ all } R \}. \quad (2.3)$$

Of course $E^q < 0$. It is well known that E^q is always finite and that the Lieb-Thirring [7] proof of stability carries through for this case [8].

Given ψ , A , and R , consider the following scaling (with $\lambda > 0$):

$$\begin{aligned} \psi(X, \underline{s}) &\rightarrow \lambda^{3N/2} \psi(\lambda X, \underline{s}), \\ A(x) &\rightarrow \lambda A(\lambda x), \\ B(x) &\rightarrow \lambda^2 B(\lambda x), \\ R &\rightarrow (1/\lambda) R. \end{aligned} \quad (2.4)$$

The various quantities scale as

$$\begin{aligned} W(\psi, R, \underline{z}) &\rightarrow \lambda W(\psi, R, \underline{z}), \\ T(\psi, A) &\rightarrow \lambda^2 T(\psi, A), \quad \tau(\psi, A) \rightarrow \lambda^2 \tau(\psi, A), \\ Q(\psi) &\rightarrow \lambda^3 Q(\psi). \end{aligned} \quad (2.5)$$

If we define

$$W(\psi, \underline{z}) = \sup_R W(\psi, R, \underline{z}), \quad (2.6)$$

then W scales as

$$W(\psi, \underline{z}) \rightarrow \lambda W(\psi, \underline{z}). \quad (2.7)$$

Note that

$$\begin{aligned} E^q(\underline{z}) &= \inf_\psi T(\psi, A) - W(\psi, \underline{z}), \\ E(\underline{z}) &= \inf_\psi \tau(\psi, A) - W(\psi, \underline{z}). \end{aligned} \quad (2.8)$$

From (2.5)–(2.7) we deduce (as in the case of the one-electron atom) that

$$4|E^q(\underline{z})| T(\psi, A) \geq W(\psi, \underline{z})^2 \geq W(\psi, R, \underline{z})^2. \quad (2.9)$$

The strategy has 7 steps.

Step 1. In [1, Lemma 3.1] a bound for τ in terms T and Q was derived (which trivially extends to N -particles). There are two cases (depending on ψ and A).

Case 1. $T(\psi, A) \geq 2Q(\psi)$. Then

$$\tau(\psi, A) \geq T(\psi, A) - Q(\psi). \quad (2.10)$$

Case 2. $T(\psi, A) \leq 2Q(\psi)$. Then

$$\tau(\psi, A) \geq \frac{1}{4} T(\psi, A)^2 / Q(\psi). \quad (2.11)$$

As will be seen, Case 1 is relevant for determining E^L while Case 2 is relevant for determining $\tilde{z}_c$.

Step 2. (This step is trivial for $K = 1$.) Pick some $\underline{z}_0 = (z_0^1, \dots, z_0^K)$ and consider the rectangle $\underline{z} < \underline{z}_0$ (which means $0 \leq z^j \leq z_0^j$, all j). For each fixed ψ and R , the minimum of $W(\psi, R, \underline{z})$ in this rectangle occurs at one of the 2^K vertices. This is proved in [2] Lemma 2.3 et. seq. From this it follows that $W(\psi, \underline{z})$, $-E^q(\underline{z})$ and $-E(\underline{z})$ are monotone nondecreasing functions of $\underline{z}$ (with the above order relation). Hence if stability holds for $\tilde{\underline{z}} = (\tilde{z}, \dots, \tilde{z})$ then it holds when all $z^j \leq \tilde{z}$.

Step 3 (Definition of $\tilde{z}_c$). Define

$$\delta(\psi, A, \underline{z}) = \frac{1}{4} T(\psi, A)^2 / Q(\psi) - W(\psi, \underline{z}). \quad (2.12)$$

The two terms of (2.12) scale the same way [see (2.5) and (2.7)], so that the infimum of $\delta(\psi, A, \underline{z})$ (over ψ and A) is either zero or $-\infty$. We define [with $\tilde{\underline{z}} \equiv (\tilde{z}, \dots, \tilde{z})$]

$$\tilde{z}_c = \sup \{ \tilde{z} \mid \delta(\psi, A, \tilde{\underline{z}}) \geq 0 \text{ for all } (\psi, A) \in \mathcal{C} \}. \quad (2.13)$$

Step 4. Suppose that $z^j < \tilde{z}_c$ for all j and let $(\psi, A) \in \mathcal{C}$ be given. If case 1, (2.10), holds then

$$\mathcal{E}(\psi, A, R, \underline{z}) \geq \frac{1}{2} T(\psi, A) - W(\psi, R, \underline{z}) \geq 2E^q(\underline{z}), \quad (2.14)$$

by scaling. If case 2 holds then $\delta(\psi, A, \underline{z}) \geq 0$. In either case $E(\underline{z})$ is finite and thus

$$\tilde{z}_c \leq z_c. \quad (2.15)$$

Step 5. We want to find a lower bound (which we call z_c^L) to $\tilde{z}_c$. A lower bound on $T(\psi, A)$ is needed and this is provided by the Lieb-Thirring estimate [9]

$$T(\psi, A)^{3/2} \geq GQ(\psi)/\alpha^2, \quad (2.16)$$

for a universal constant $G = 1.28$, explicated in (3.8). This leads to the bound

$$\delta(\psi, A, \underline{z}) \geq \frac{1}{4} (G\alpha^{-2}) T(\psi, A)^{1/2} - W(\psi, \underline{z}). \quad (2.17)$$

Combining this with the bound (2.9) [and the trivial fact that we need only consider $W(\psi, \underline{z}) \geq 0$] we see that $\delta(\psi, A, \underline{z}) \geq 0$ if

$$|E^q(\underline{z})| \leq (G/8\alpha^2)^2. \quad (2.18)$$

By (2.13)

$$\tilde{z}_c \geq z_c^L \equiv \sup \{ z \mid |E^q(\underline{z})| \leq (G/8\alpha^2)^2 \}, \quad (2.19)$$

[$\underline{z}$ means $(z, \dots, z)$]. The monotonicity given in Step 2 has been used.

Step 6 (Bound on the energy). Suppose that $z^j \leq z_c^L$ for all j . Let $(\psi, A) \in \mathcal{C}$ be given. Case 2 is irrelevant since $\delta(\psi, A, \underline{z}) \geq 0$ by definition. Therefore a lower bound, $E^L(\underline{z})$, to $E(\underline{z})$ can be obtained by the following minimization problem:

$$E^L(\underline{z}) = \min(T - Q - W), \quad (2.20)$$

under the conditions

$$T \geq 2Q, T \geq (GQ/\alpha^2)^{2/3}, T \geq W^2/4|E^q(\underline{z})|. \quad (2.21)$$

This algebraic problem is solved in Appendix B of [1] and the result is

$$E(\underline{z}) \geq E^L(\underline{z}) = E^q(\underline{z})f(\gamma), \quad (2.22)$$

$$f(\gamma) \equiv \frac{4}{3}\gamma^{-2}\{3\gamma - 2 + 2(1 - \gamma)^{3/2}\}, \quad (2.23)$$

$$\gamma \equiv 6|E^q(\underline{z})|^{1/2}\alpha^2/G. \quad (2.24)$$

Equation (2.22) gives E^L as the exact E^q times a correction factor, f , which depends on γ , where γ is proportional to $|E^q|^{1/2}$. Two things should be noted: By the definition (2.19),

$$\gamma \leq 3/4, \quad (2.25)$$

when $z^j < z_c^L$ (all j). Second, the function f is monotone increasing in γ on $[0, 1]$.

Step 7. To utilize (2.19) and (2.22) we require a bound on $E^q(\underline{z})$. Let

$$E_L^q(\underline{z}) \leq E^q(\underline{z}) < 0, \quad (2.26)$$

be any lower bound to E^q . Inserting $E_L^q(\underline{z})$ in (2.19) will give a lower bound to $z_c^L \leq z_c$. Inserting $E_L^q(\underline{z})$ in (2.24) and then inserting this γ in (2.23) and (2.22) will (assuming that $\gamma \leq 1$) give a lower bound to E^L . In cases A and B we can get an effective $E_L^q(\underline{z})$ which is independent of N and K . The former uses the Lieb-Thirring technique [7] together with a novel bound on the Coulomb energy. This is done in Sect. III. Case B is controlled by relating it to a relativistic problem solved in [2]; this is done in Sect. IV.

Remark. In case B we deal with only one electron. Given this restriction on N , (2.16) holds with a larger value of G , namely $G = 3.83$. This larger G can be used in Steps 5–7.

III. The Many-Electron Atom

Our first task is to prove the kinetic energy estimate (2.16). Consider the single-particle Schrödinger operator $h = (p - A)^2 - V(x)$, where $V(x) \geq 0$ and consider also the N -particle operator $H = \sum_j h_j$. The q spin state fermionic ground state energy of H , E , satisfies $E \geq q \sum_i e_i$, where the e_i are the negative eigenvalues of h . ($q = 2$ in our case.) We have that

$$\sum_i e_i \geq -|e_1|^{1/2} \sum_i |e_i|^{1/2}, \quad (3.1)$$

where e_1 is the ground state energy. In [1, (3.19)] we quoted a result of [9] that

$$|e_1|^{1/2} \leq L_{\frac{1}{2}, 3}^1 \|V\|_2^2, \quad (3.2)$$

where $L_{1/2, 3}^1 = 0.0135$ to three significant figures.

In [9] it is also shown that

$$\sum_i |e_i|^{1/2} \leq L_{\frac{1}{2}, 3} \|V\|_2^2. \quad (3.3)$$

Strictly speaking, (3.3) was shown only for $A=0$ in [9] and it is not known whether the (unknown) *sharp* constant L in (3.3) occurs for $A=0$. However, as pointed out in [8, 11], the L actually obtained in [9] holds for all A . The L obtained by using the method of [12] also holds for all A (see [11] for a discussion of the Ito-Nelson integral). The latter method gives a better value for L and the numerical computation is most clearly explained in [10, Eqs. (46)–(51)]. In the notation of [10], we take $a=0.61$ exactly and $b=3.6807$. Then (3.3) holds with

$$L_{\frac{1}{2}, 3} = b(4\pi)^{-3/2} \Gamma(\frac{3}{2})^{\frac{1}{2}} a^{-1} = 0.060021, \quad (3.4)$$

to 5 figures. Thus,

$$\sum_i |e_i| \leq L_{\frac{1}{2}, 3}^1 L_{\frac{1}{2}, 3} \|V\|_2^4 \leq (0.000810) \|V\|_2^4. \quad (3.5)$$

Now take $V(x) = c\varrho_\psi(x)$, where ϱ_ψ is given by (2.1). Then

$$T(\psi, A) - c \int \varrho_\psi^2 = (\psi, H\psi) \geq -q \sum_i |e_i|. \quad (3.6)$$

Using (3.5) and (3.6), with $c^{-3} = 4qL_{\frac{1}{2}, 3}^1 L_{\frac{1}{2}, 3} \int \varrho_\psi^2$, we obtain

$$T(\psi, A) \geq \frac{3}{4} (4qL_{\frac{1}{2}, 3}^1 L_{\frac{1}{2}, 3})^{-1/3} \{ \int \varrho_\psi^2 \}^{2/3} \geq (4.02) \{ \int \varrho_\psi^2 \}^{2/3}, \quad (3.7)$$

for $q=2$. Thus, (2.16) holds [recalling (2.2)] with

$$G = 8.07/2\pi = 1.28. \quad (3.8)$$

Our second task is to find a lower bound for $E^q(\underline{z})$, given by (2.3). Again we use an inequality derived in [7, 9], but with a better constant derived in [10, Eq. (52)]:

$$T(\psi, A) \geq (2.7709) \int \varrho_\psi(x)^{5/3} dx. \quad (3.9)$$

The second term in V , (1.5), is absent since there is only one nucleus, located at $R=0$. The third term contributes the following to W :

$$W_3(\psi, z) = z \int \varrho_\psi(x) |x|^{-1} dx. \quad (3.10)$$

The first term in V (call its contribution W_1) requires some elaboration. For $x, y \in \mathbb{R}^3$ and $R > 0$,

$$|x - y|^{-1} \geq \{|x| + |y|\}^{-1} \geq \frac{1}{2} R f(x) f(y), \quad (3.11)$$

$$\begin{aligned} f(x) &= 1/|x| & \text{if } |x| \geq R, \\ &= 0 & \text{if } |x| < R. \end{aligned} \quad (3.12)$$

Using (3.11) and the positivity of $|\psi|^2$ and $|x_i - x_j|^{-1}$, we have, for any $0 \leq \sigma \leq 1$,

$$\begin{aligned} W_1(\psi, z) &\leq -\frac{1}{4} R \sigma \left(\psi, \left\{ \left[\sum_{i=1}^N f(x_i) \right]^2 - \sum_{i=1}^N f(x_i)^2 \right\} \psi \right) \\ &\leq -\frac{1}{4} R \sigma \{ [\int \varrho_\psi f]^2 - \int \varrho_\psi f^2 \}, \end{aligned} \quad (3.13)$$

since $\langle (\sum f)^2 \rangle \geq \langle \sum f^2 \rangle$.

Combining (3.9), (3.10), (3.13),

$$\begin{aligned} T(\psi, A) - W(\psi, z) &\equiv \mathcal{E}^q(\psi, A) \geq \mathcal{E}_{R, \sigma}(\varrho_\psi, z) \\ &\equiv (2.7709) \int \varrho_\psi^{5/3} - z \int \varrho_\psi |x|^{-1} + \frac{1}{4} R \sigma [\int \varrho_\psi f]^2 \\ &\quad - \frac{1}{4} R \sigma \int \varrho_\psi f^2. \end{aligned} \quad (3.14)$$

Therefore,

$$E^q(z) \geq \sup_{0 \leq \sigma \leq 1} \sup_{R > 0} \inf_{\varrho} \mathcal{E}_{R, \sigma}(\varrho, z). \quad (3.15)$$

We could, of course, impose the extra condition $\int \varrho = N$ in (3.15) but, as we desire an N -independent bound for E^q , we forego this.

First minimize (3.14) with respect to $\varrho(x)$ for $|x| \leq R$. Only the first two terms are relevant in this region. Define $\Gamma = (5/3)(2.7709)$. Then $\Gamma \varrho^{2/3}(x) = z/|x|$. The first two terms contribute (for $|x| < R$)

$$-z^{5/2} \Gamma^{-3/2} (16\pi/5) R^{1/2}. \quad (3.16)$$

Next we consider the contributions for $|x| > R$. Here we merely omit the $\varrho^{5/3}$ term and we use $R|x|^{-2} \leq |x|^{-1}$ in the last term. Let $Y \equiv \int_{|x| > R} \varrho(x) |x|^{-1} dx$. Then the sum of the last three terms is not less than the minimum (with respect to Y) of $-(z + \frac{1}{4}\sigma)Y + \frac{1}{4}R\sigma Y^2$. This minimum is

$$-(z + \frac{1}{4}\sigma)^2 / R\sigma. \quad (3.17)$$

The maximum of this with respect to $\sigma \in [0, 1]$ is

$$-M(z)/R, \quad (3.18)$$

$$\begin{aligned} M(z) &= z && \text{if } z \leq 1/4, \\ &= (z + \frac{1}{4})^2 && \text{if } z \geq 1/4. \end{aligned} \quad (3.19)$$

Adding (3.16) and (3.18) and then maximizing with respect to $R > 0$ gives

$$\begin{aligned} E^q(z) &\geq -3z^{5/3} \Gamma^{-1} (8\pi/5)^{2/3} M(z)^{1/3} \\ &= -(1.9062) z^{5/3} M(z)^{1/3}. \end{aligned} \quad (3.20)$$

As we shall be primarily interested in $z > 1/4$, the little exercise with σ is academic; it was done merely to demonstrate a z^2 (instead of $z^{5/3}$) bound when $z \leq 1/4$.

With these results we can now bound z_c , see (2.19) and E^L , see (2.22). Since z_c will be large, let us use the bound $z^{5/3} M(z)^{1/3} \leq (z + \frac{1}{4})^{7/3}$ for all $z > 0$. Then, from (2.19)

$$z_c \geq \tilde{z}_c \geq z_c^L \geq -\frac{1}{4} + (0.158)\alpha^{-12/7} \geq 720. \quad (3.21)$$

This bound (720) is about 25 times smaller than the z_c^L obtained in [1] for the one-electron atom. It is about 290 times less than the *upper* bound on z_c obtained in [13], see also [1, (3.24)]. This upper bound ($z_c \leq 208,000$) also holds, of course, for the full problem with K nuclei and N electrons.

The lower bound (2.22) on the energy is

$$E^L = E^q(z) f(\gamma) \quad (3.22)$$

and, using (3.20),

$$\begin{aligned} \gamma &\leq 6\alpha^2(1.9062)^{1/2} (z + \tfrac{1}{4})^{7/6} / G \\ &\leq (6.47) \alpha^2 (z + \tfrac{1}{4})^{7/6} \\ &\leq (0.000345) (z + \tfrac{1}{4})^{7/6}. \end{aligned} \quad (3.23)$$

As an illustration, take $z = 100$. By (2.23) the fractional change in the energy, $f(\gamma) - 1$, is less than 0.013, which is about $1\frac{1}{2}\%$.

IV. The One-Electron Molecule

Our first task is to find a lower bound to E^q in (2.3) with

$$V(x, R, z) = - \sum_{j=1}^K z^j |x - R_j|^{-1} + \sum_{i < j} z^i z^j |R_i - R_j|^{-1}. \quad (4.1)$$

Since $N = 1$, we can use the diamagnetic inequality (see [1]): $T(\psi, A) \geq T(|\psi|, 0) \equiv T(\psi) = \|\nabla|\psi|\|_2^2$, and hence can assume that ψ is real and positive and $A = 0$. Define

$$\bar{V}(x, R) = -(2/\pi) \sum_{j=1}^K |x - R_j|^{-1} + (12/\pi) \sum_{i < j} |R_i - R_j|^{-1}. \quad (4.2)$$

It is proved in [2, Proposition 2.2] that for all $\psi \in L^2$, $(-\Delta)^{1/4} \psi \in L^2$ and all R ,

$$(\psi, (-\Delta)^{1/2} \psi) \geq -(\psi, \bar{V} \psi). \quad (4.3)$$

We also have the fact (Schwarz inequality) that

$$\|\nabla \psi\|_2^2 \geq (\psi, (-\Delta)^{1/2} \psi)^2, \quad (4.4)$$

when $\|\psi\|_2 = 1$.

Given z , define

$$Z = \max(z^1, \dots, z^K) \quad \text{and} \quad \underline{Z} = (Z, \dots, Z). \quad (4.5)$$

As shown in Step 2,

$$E^q(z) \geq E^q(\underline{Z}). \quad (4.6)$$

Suppose that $Z \geq 6$. Then

$$(\pi Z/2) \bar{V}(x, R) \leq V(x, R, \underline{Z}). \quad (4.7)$$

Combining (4.3), (4.4), (4.7) and with $t = (\psi, (-\Delta)^{1/2}\psi)$

$$E^q(z) \geq \inf_t \{t^2 - (\pi Z/2)t\} = -(\pi Z/4)^2. \quad (4.8)$$

[Note: When $K = 1$, the exact result is $-(Z/2)^2$.]

By monotonicity (4.6), when $Z < 6$

$$E^q(Z) \geq E^q(6) \geq -(3\pi/2)^2. \quad (4.9)$$

Combining (4.6), (4.8), (4.9) we obtain for all z

$$|E^q(z)|^{1/2} \leq (\pi/4) \max\{6, z^1, \dots, z^K\}. \quad (4.10)$$

Turning now to (2.19) and using (4.10) we have that

$$z_c^L \geq \sup \left\{ z \left| \frac{\pi}{4} \max(6, z) \leq G/8\alpha^2 \right. \right\}. \quad (4.11)$$

As remarked at the end of Sect. II, since $N = 1$ we are entitled to replace $L_{\frac{1}{2},3}$ by $L_{\frac{1}{4},3}^1$ in (3.7), (3.8), and (2.16). Thus,

$$G = 3.83, \quad (4.12)$$

in our case.

Suppose that

$$\alpha^2 \leq \alpha_c^2 \equiv G/(12\pi) = 0.102. \quad (4.13)$$

Then, from (4.11)

$$z_c^L \geq G/(2\pi\alpha^2) = 0.609\alpha^{-2} > 11,400. \quad (4.14)$$

(This number, 11,400, compares favorably with 17,900 obtained in [1] for $K = 1$.) In the opposite case [(4.13) is violated], the set of z 's in (4.13) is empty and our method gives no bound at all on $E(z)$ for $z \neq 0$. Thus, our method requires *two* conditions for stability

$$(i) \quad \alpha^2 z^j \leq 0.609 \quad \text{for all } j, \quad (4.15)$$

$$(ii) \quad \alpha \leq \alpha_c = (0.102)^{1/2} = 0.319. \quad (4.16)$$

One can question whether the condition (4.16) on α is an artifact of our method or whether there really is an α_c (which will, of course, be greater than 0.319 – but finite). The second alternative is correct as we now prove.

Lemma. Suppose that

$$\alpha > 6.67, \quad (4.17)$$

then for every $z = (z, \dots, z)$ with $z > 0$ there is a K such that $E(z) = -\infty$.

Remark. The right side of (4.17) is not the best bound that can be obtained by the following method.

Proof. In [1] we showed that $E = -\infty$ when $K = 1$ if

$$z\alpha^2 > \inf \{ \int B^2 \} \{ 8\pi(\psi, |x|^{-1}\psi) \}^{-1} \equiv P, \quad (4.18)$$

where (ψ, A) runs over $\mathcal{F} = \{(\psi, A) \in \mathcal{C} \mid \sigma \cdot (p - A)\psi = 0\}$. $\mathcal{F}$ is not empty [13]. By taking a particular example, one finds $P \leq 9\pi^2/8 = 11.10$. Therefore, if $\alpha^2 > P$, we can take $K = 1$ and achieve instability for all $z \geq 1$. Using the above bound, this is also achieved for $z \geq 1$ if $\alpha > 3.34$.

Next, to investigate $z < 1$, take any $(\psi, A) \in \mathcal{F}$, whence

$$\mathcal{E}(\psi, A, R, z) = \varepsilon \int B^2 + \int \varrho_\psi(x) V(x, R, z) dx, \quad (4.19)$$

with $\varrho_\psi(x) = \langle \psi, \psi \rangle(x)$. We want to show that for suitable α and K , $\mathcal{E}$ is negative for some R . [If it is negative then, by the scaling (2.4), $\mathcal{E}$ can be made arbitrarily negative.] To show this, it suffices to average $\mathcal{E}$ with some probability density $F(R^1, \dots, R^K)$, $\int F d^K R = 1$, and to show that $\langle \mathcal{E} \rangle \equiv \int \mathcal{E} F d^K R < 0$. Take $F = \varrho_\psi(R^1) \dots \varrho_\psi(R^K)$. The result is

$$\langle \mathcal{E} \rangle = \varepsilon \int B^2 - \frac{1}{2} z K [2 - z(K-1)] I(\varrho_\psi), \quad (4.20)$$

$$I(\varrho) = \int \int \varrho(x) \varrho(y) |x - y|^{-1} dx dy. \quad (4.21)$$

Choose K to be the smallest integer closest to $\frac{1}{2} + 1/z$. Then $zK = (z/2) + 1 + \mu$ with $|\mu| \leq \frac{1}{2}z$ and $zK[2 - z(K-1)] = [1 + (z/2)]^2 - \mu^2 \geq 1 + z > 1$. Therefore, if

$$\alpha^2 > (4\pi)^{-1} \inf_{\mathcal{F}} \int B^2 / I(\varrho), \quad (4.22)$$

instability occurs for all $0 < z < 1$.

For the particular example in [13] quoted above, one has

$$|B(x)| = 12(1 + |x|^2)^{-2}, \quad \varrho(x) = [\pi(1 + |x|^2)]^{-2},$$

and one computes

$$\int B^2 = 18\pi^2, \quad I(\varrho) = 1/\pi. \quad (4.23)$$

Therefore, if $\alpha > 3 \cdot 2^{-1/2} \pi = 6.67$, instability also occurs for all $z < 1$. $\square$

Acknowledgements. It is a pleasure to thank J. Fröhlich and H.-T. Yau for helpful discussions.

References

1. Fröhlich, J., Lieb, E.H., Loss, M.: Stability of Coulomb systems with magnetic fields. I. The one-electron atom. *Commun. Math. Phys.* **104**, 251–270 (1986)
2. Daubechies, I., Lieb, E.H.: One electron relativistic molecules with Coulomb interaction. *Commun. Math. Phys.* **90**, 497–510 (1983)
3. Conlon, J.: The ground state energy of a classical gas. *Commun. Math. Phys.* **94**, 439–458 (1984)
4. Fefferman, C., de la Llave, R.: Relativistic stability of matter I. *Revista Iberoamericana* (to appear)
5. Ni, G., Wang, Y.: Vacuum instability and the critical value of the coupling parameter in scalar QED. *Phys. Rev. D* **27**, 969–975 (1983)
6. Finger, J., Horn, D., Mandula, J.E.: Quark condensation in quantum chromodynamics. *Phys. Rev. D* **20**, 3253–3272 (1979)
7. Lieb, E.H., Thirring, W.: Bound for the Kinetic energy of fermions which proves the stability of matter. *Phys. Rev. Lett.* **35**, 687 (1975). Errata **35**, 1116 (1975)
8. Avron, J., Herbst, I., Simon, B.: Schrödinger operators with magnetic fields. I. General interactions. *Duke Math. J.* **45**, 847–883 (1978)

9. Lieb, E.H., Thirring, W.: Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In: *Studies in Mathematical Physics, Essays in honor of Valentine Bargmann*, Lieb, E.H., Simon, B., Wightman, A.S. (eds.). Princeton, NJ: Princeton University Press 1976
10. Lieb, E.H.: On characteristic exponents in turbulence. *Commun. Math. Phys.* **92**, 473–480 (1984)
11. Simon, B.: *Functional integration and quantum physics*. New York: Academic Press 1979
12. Lieb, E.H.: The number of bound states of one-body Schrödinger operators and the Weyl problem. *Proc. Am. Math. Soc. Symp. Pure Math.* **36**, 241–252 (1980)
13. Loss, M., Yau, H.-T.: Stability of Coulomb systems with magnetic fields. III. Zero energy bound states of the Pauli operator. *Commun. Math. Phys.* **104**, 283–290 (1986)

Communicated by A. Jaffe

Received October 10, 1985

PHYSICAL REVIEW LETTERS

VOLUME 75

7 AUGUST 1995

NUMBER 6

Stability of Matter in Magnetic Fields

Elliott H. Lieb,^{1,2} Michael Loss,³ and Jan Philip Solovej²¹*Department of Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544*²*Department of Mathematics, Fine Hall, Princeton University, Princeton, New Jersey 08544*³*School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332*

(Received 12 April 1995)

In the presence of arbitrarily large magnetic fields, matter composed of electrons and nuclei was known to be unstable if α or Z is too large. Here we prove that matter is stable if $\alpha < 0.06$ and $Z\alpha^2 < 0.04$.

PACS numbers: 03.65.-w, 11.10.-z, 12.20.Ds, 31.10.+z

One of the remaining unsolved problems connected with the stability of matter is the inclusion of arbitrary magnetic fields. The model is a caricature of QED, which invites speculations about stability of QED for large fine structure constant α , but that is not our focus here and we refer to [1] for a discussion of these and related matters. The Hamiltonian for N electrons and K fixed nuclei of charge Ze with magnetic field $\mathbf{B}(x) = \nabla \times \mathbf{A}(x)$, including the field energy, $\varepsilon \int B^2$, is

$$H = \sum_{i=1}^N \mathcal{T}_i + V_C + \varepsilon \int B(x)^2 d^3x, \quad (1)$$

where $\mathcal{T} \equiv [\boldsymbol{\sigma} \cdot (\mathbf{p} + \mathbf{A})]^2 = (\mathbf{p} + \mathbf{A})^2 + \boldsymbol{\sigma} \cdot \mathbf{B}$ is the Pauli operator. The Coulomb energy is

$$V_C = -Z \sum_{i=1}^N \sum_{j=1}^K |x_i - R_j|^{-1} + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} + Z^2 \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1}, \quad (2)$$

with R_j being the coordinates of the nuclei and x_i the electron coordinates. The energy unit is $4 \text{ Ry} = 2mc^2\alpha^2$, $\alpha = e^2/\hbar c$, length unit $= \hbar^2/2me^2$, and $\varepsilon = (8\pi\alpha^2)^{-1}$. Notice that α appears in (1) only through ε .

The negative particles are necessarily spin 1/2 fermions which, for mathematical generality, we assume to exist in $q/2$ flavors (e.g., $q = 6$, corresponding to three leptons). The ground state energy is denoted by E .

Starting with the 1967 pioneering work of Dyson and Lenard we now understand stability for arbitrarily many electrons and nuclei, with $\mathbf{B} = 0$, in the context

of the nonrelativistic Schrödinger equation. Later it was extended to the “relativistic” Schrödinger equation in which $\mathbf{p}^2/2m$ is replaced by $(c^2\mathbf{p}^2 + m^2c^4)^{1/2}$ (see [2] for a review). These proofs also hold with the inclusion of a magnetic field coupled to the orbital motion of the electrons, i.e., $\mathbf{p} \rightarrow \mathbf{p} + \mathbf{A}$, but no Zeeman $\boldsymbol{\sigma} \cdot \mathbf{B}$ term.

Stability of matter has two meanings: (i) E is finite for arbitrary N and K ; (ii) $E \geq -C_1(N + K)$ for some constant C_1 independent of N , K , and R_j . (ii) obviously implies (i), and it holds in the nonrelativistic case. In the relativistic case, (i) actually implies (ii) (see [3]) but (i) requires two conditions: $Z\alpha \leq C_2$ and $\alpha \leq C_3$ with C_2 and C_3 being universal constants, the best available values being in [3], Theorems 1 and 2. The inclusion of $\mathbf{B}$ changes E , but the point is that while C_1, C_2, C_3 depend on q , they can be chosen to be independent of $\mathbf{B}$.

The situation changes dramatically when the magnetic moments of the electrons are allowed to interact with the magnetic field via the $\boldsymbol{\sigma} \cdot \mathbf{B}$ term, as in (1). The reason for this is simple: The Pauli operator $\mathcal{T}$ is non-negative, but it is much weaker than $(\mathbf{p} + \mathbf{A})^2$. Indeed, it can even have square integrable zero modes [4], $\mathcal{T}\psi = 0$, for suitable $\mathbf{A}(x)$, which cause instability for large $Z\alpha^2$.

It is known [5] that without the field energy term $\varepsilon \int B^2$ in (1) arbitrarily large $\mathbf{B}$ fields can cause arbitrarily negative energies E even for hydrogen. The field energy, hopefully, stabilizes the situation, and our goal is to show that E is finite for (1), even after minimizing over all possible $\mathbf{B}$ fields and all possible R_j .

One of our results on magnetic stability is as follows.

Theorem 1: The ground state energy of H satisfies

$$E \geq -2.6q^{2/3} \max\{Q(Z)^2, Q(5.7q)^2\} N^{1/3} K^{2/3}, \quad (3)$$

with $Q(t) \equiv t + \sqrt{2t} + 2.2$, provided that

$$qZ\alpha^2 \leq 0.082 \quad \text{and} \quad q\alpha \leq 0.12. \quad (4)$$

In (2) all the nuclear charges are set equal to Z . As far as stability is concerned this is no restriction [6], since the energy is concave in each charge Z_j and hence stability holds in the “cube” $\{0 \leq Z_j \leq Z\}_{j=1}^K$ if it holds when all $Z_j = Z$. It also follows from this that E is a decreasing function of Z . Moreover, since $\varepsilon \propto \alpha^{-2}$, it follows that E is a decreasing function of α , a fact that will be important later. The form of (3) is the best possible for $Z \geq 1$, as we know from other studies [2].

Our actual condition for stability given after (18) is rather complicated, but very much more general than (4)—which is only representative. The results after (18) show, e.g., that when $\alpha = 1/137$ and $q = 2$, Z can be as large as 1050. The large values of Z and α are important because the comfortable distance of the critical values from the physical values $Z \leq 92$, $\alpha = 1/137$ implies that the effect studied here is merely a small perturbation of the usual $B = 0$ case.

Our proof of Theorem 1 will require a new technique—a running energy-scale renormalization of $\mathcal{T}$. A byproduct of this is a Lieb-Thirring type inequality for $\mathcal{T}$:

Theorem 2: If $\varepsilon_1 \leq \varepsilon_2, \dots < 0$ are the negative eigenvalues of $\mathcal{T} - U$, for a potential $-U(x) \leq 0$ then

$$\sum |\varepsilon_i| \leq a_\gamma \int U(x)^{5/2} d^3x + b_\gamma \left(\int B(x)^2 d^3x \right)^{3/4} \left(\int U(x)^4 d^3x \right)^{1/4} \quad (5)$$

for all $0 < \gamma < 1$, where $a_\gamma = (2^{3/2}/5)(1 - \gamma)^{-1} L_3$ and $b_\gamma = 3^{1/4} 2^{-9/4} \pi \gamma^{-3/8} (1 - \gamma)^{-5/8} L_3$. We can take L_3 , defined below, to be 0.1156.

More generally, the second term in (5) can be replaced by $(\int B^{3q/2})^{1/q} (\int U^p)^{1/p}$, where $p^{-1} + q^{-1} = 1$.

The investigation of this problem started in [1,7] where type (ii) stability was proved (for suitable Z, α and $q = 2$) for $K = 1$ and arbitrary N [if $(Z + 1/4)\alpha^{12/7} \leq 0.15$] or $N = 1$ and arbitrary K (if $Z\alpha^2 \leq 0.6$ and $\alpha \leq 0.3$). The problem for general N and K was open for nine years, and we present a surprisingly simple solution here.

The bounds in (4) on $Z\alpha^2$ and α are not artifacts. It is shown in [1] and [4] that the zero modes cause $E = -\infty$ when $Z\alpha^2 > 11.11$ for the “hydrogenic” atom, i.e., a single spin 1/2 particle and one nucleus. If the number of nuclei is arbitrary, it is shown in [7] that there is collapse if $\alpha > 6.67$, no matter how small Z is. *Magnetic stability, like relativistic stability, implies a (Z -independent) bound on α .*

986

Prior to our work a proof of type (ii) stability for (1) with $Z = 1$, $q = 2$, and some sufficiently small α was announced (unpublished) by C. Fefferman and sketched to one of us. Our proof is unrelated to his, considerably simpler and, more importantly, gives physically realistic constants.

We begin our analysis with the observation that length scaling considerations suggest that the key to understanding the stability problem is somehow to replace $\mathcal{T}$, on each energy scale e , by $\mu\mathcal{T}/e$, where μ is a fixed energy but e is variable. On energy scales $e > \mu$ we can use the fact that $\mathcal{T} > 0$ to replace $\mathcal{T}$ by $\mu\mathcal{T}/e$ without spoiling lower bounds. It might seem odd to replace $\mathcal{T}$ by something smaller, but what is really happening is that $\sigma \cdot B$ is being partially controlled by $[1 - \mu e^{-1}](p + A)^2$. The idea of replacing $\mathcal{T}$ by a fraction of $\mathcal{T}$ was also used in [1], but no energy dependence was used there.

We shall illustrate this concept by three calculations. The first, (A), will establish magnetic stability by relating it to the stability of relativistic matter (see [3,6,8,9]). The second, (B), will be the proof of Theorem 2. The third, (C), will use essential parts of the second calculation and an electrostatic inequality proved in [3] to prove magnetic stability without resorting to relativistic stability.

(A) *Magnetic stability from relativistic stability.*—We use stability of relativistic matter in the form proved in [3]. From the corollary of Theorem 1 in [3] with $\beta = 0.5$ we have, for any $0 < q\kappa \leq 0.032$ and $Z\kappa \leq 1/\pi$,

$$\sum_{i=1}^N |p_i + A_i| + \kappa V_C \geq 0. \quad (6)$$

(Although Theorem 1 in [3] was stated only for $|p|$, it holds for $|p + A|$ because it relies only on the magnitude of the resolvent, which only gets smaller when A is not zero. That is, $\| |p + A|^{-s}(x, y) \| \leq \| |p|^{-s}(x, y) \|$ for each $s > 0$ and x, y in $\mathbb{R}^3$. This follows at once from a similar bound on the heat kernel $\{\exp[-t(p + A)^2]\}(x, y)$ which, in turn, follows from its representation as a path integral. This was pointed out in [5,10]. Only the resolvent powers $|p + A|^{-s}$ enter the proof of Theorem 1 in [3].)

Using (6), H is bounded below by $\bar{H} = \sum_{i=1}^N h_i + \varepsilon \int B^2$, where h is the one-body operator $h = \mathcal{T} - \kappa^{-1}|p + A|$. Thus E is bounded below by $\varepsilon \int B^2 + \bar{E}_N$, where $\bar{E}_N = q \sum_{j=1}^{[N/q]} \varepsilon_j$ and $\varepsilon_1 \leq \varepsilon_2 \leq \dots$ are the eigenvalues of h . For $e > 0$, let $N_{-e}(h)$ be the number of eigenvalues of h less than or equal to $-e$. Choose $\mu > 0$ and note that

$$\bar{E}_N \geq -N\mu - q \int_{\mu}^{\infty} N_{-e}(h) de. \quad (7)$$

The crucial step in our proof is noting that the positivity of the operator $\mathcal{T}$ implies that $\mathcal{T} \geq \mu\mathcal{T}/e$ when $e \geq \mu$. Thus $\mathcal{T} \geq \mu e^{-1} \mathcal{T} \geq \mu e^{-1}(p + A)^2 - \mu e^{-1} B(x)$

when $e \geq \mu$. By Schwarz's inequality, $\kappa^{-1}|p + A| \leq (1/3)e^{-1}\kappa^{-2}(p + A)^2 + 3e/4$ and hence if we set $\mu = (4/3)\kappa^{-2}$ we obtain

$$h \geq e^{-1}\kappa^{-2}(p + A)^2 - \frac{4}{3}e^{-1}\kappa^{-2}B(x) - 3e/4 \equiv h_e.$$

Thus $N_{-e}(h) \leq N_{-e}(h_e)$, and this can be estimated by the Cwikel-Lieb-Rozenblum (CLR) bound [11], i.e., $N_{-e}[(p + A)^2 - U(x)] \leq L_3 \int [U(x) - e]_+^{3/2} d^3x$, where $[a]_+ \equiv \max(a, 0)$ and $L_3 = 0.1156$. In our case

$$N_{-e}(h_e) \leq L_3 \int \left[\frac{4B(x)}{3} - \frac{e^2\kappa^2}{4} \right]_+^{3/2} d^3x. \quad (8)$$

Inserting this bound in (7), a simple calculation yields

$$\bar{E}_N \geq -N\mu - (2\pi/3)q\kappa^{-1}L_3 \int B(x)^2 d^3x.$$

We choose κ so that the field energy terms are non-negative, i.e., $\kappa \geq (16\pi^2/3)L_3\alpha^2q = 6.1\alpha^2q$. We conclude, by (6), that magnetic stability holds if

$$q\alpha \leq 0.071 \quad \text{and} \quad qZ\alpha^2 \leq 0.052. \quad (9)$$

For $q = 2$, the first condition is $\alpha \leq 1/28$. For $q = 2$ and $\alpha = 1/137$, stability occurs if $Z \leq 490$.

Assuming (9) holds, we then use (6) and choose $\kappa = \min\{0.0315q^{-1}, (\pi Z)^{-1}\}$. Our lower bound on the ground state energy per electron, by this method, is then $-\mu = -(4/3)\kappa^{-2} = -\max\{1345q^2, 13.2Z^2\}$.

Remark: We used the CLR bound in (8). Since the derivation of this bound is not elementary, the reader might wish to use an easier to derive bound—at the cost of worsening the final constants. A useful substitute is

$$N_{-e} \leq 0.1054e^{-1/4} \int [U(x) - e/2]_+^{7/4} d^3x$$

(plus an increased μ), which is in (2.8) of [12] and which can be derived by means originally employed for the Lieb-Thirring inequality. This same remark also applies to our other calculations below.

(B) *The Lieb-Thirring inequality.*—As before we note that $\sum \varepsilon_i = -\int_0^\infty N_{-e}(\mathcal{T} - U) de$. We write $\int_0^\infty = \int_0^\mu + \int_\mu^\infty$. The parameter μ will be optimized below. Noting that $\mathcal{T} \geq (p + A)^2 - B(x)$ and applying the CLR bound in the same fashion as before to $\int_0^\mu$ yields

$$L_3 \int_0^\mu \int [B(x) + U(x) - e]_+^{3/2} d^3x de. \quad (10)$$

In $\int_\mu^\infty$ we replace $\mathcal{T}$ by the lower bound $\mu e^{-1}[(p + A)^2 - B(x)]$ and obtain $N_{-e}(\mathcal{T} - U) \leq N_{-e}(\mu e^{-1}[(p + A)^2 - B] - U)$. A further application of the CLR inequality yields the bound on $\int_\mu^\infty$

$$L_e \int_\mu^\infty \int [B(x) + (e/\mu)U(x) - e^2/\mu]_+^{3/2} dx de. \quad (11)$$

It is easy to see that for any $0 < \gamma < 1$ the integrand in (10) is bounded above by

$$\sqrt{2}([B(x) - \gamma e^2/\mu]_+^{3/2} + [U(x) - (1 - \gamma)e]_+^{3/2}).$$

Treating the integrand in (11) in a similar fashion and combining the inequalities we find

$$\begin{aligned} \sum |\varepsilon_i| \leq & \sqrt{2}L_3 \int \left\{ \int_0^\infty [B(x) - \gamma e^2/\mu]_+^{3/2} de \right. \\ & + \int_0^\mu [U(x) - (1 - \gamma)e]_+^{3/2} de \\ & \left. + \int_\mu^\infty [(e/\mu)U(x) - (1 - \gamma)e^2/\mu]_+^{3/2} de \right\} d^3x. \end{aligned}$$

After extending the last two integrals to $\int_0^\infty$, a straightforward computation yields

$$\begin{aligned} \sum |\varepsilon_i| \leq & \sqrt{2}L_3 \int \left\{ \frac{2}{5(1 - \gamma)} U(x)^{5/2} + \frac{3\pi\mu^{1/2}}{16\gamma^{1/2}} B(x)^2 \right. \\ & \left. + \frac{3\pi}{128} \mu^{-3/2}(1 - \gamma)^{-5/2} U(x)^4 \right\} d^3x. \end{aligned}$$

Optimizing over μ yields (5).

To prove the more general form of (5), replace μe^{-1} by $(\mu e^{-1})^s$, where $s = 2p/3 - 5/3$.

(C) *Proof of Theorem 1.*—We turn now to our third illustration of the concept of running energy scale and prove the stability directly, not relating it to the relativistic problem. By this method we get the correct dependence of the ground state energy on Z and also somewhat better critical constants than in (9).

Following [3] we first replace the Coulomb potential by a single particle potential in (12) below. We break up $\mathbf{R}^3$ into Voronoi cells defined by the nuclear locations, i.e., $\Gamma_j = \{x : |x - R_j| \leq |x - R_k| \text{ for all } k\}$ is the j th Voronoi cell. Each Γ_j contains a ball centered at R_j with radius $D_j = \min\{|R_j - R_k| : j \neq k\}/2$.

The following bound on V_C is proved in [3]: Choose some $0 < \lambda < 1$. Then

$$V_C \geq -\sum_{i=1}^N W(x_i) + \sum_{j=1}^K \frac{Z^2}{8D_j}, \quad (12)$$

where $W(x) = Z|x - R_j|^{-1} + F_j(x)$ for $x \in \Gamma_j$ with $F_j(x)$ defined by

$$\begin{aligned} (2D_j)^{-1}(1 - D_j^{-2}|x - R_j|^2)^{-1} & \text{ for } |x - R_j| \leq \lambda D_j, \\ (\sqrt{2Z} + 1/2)|x - R_j|^{-1} & \text{ for } |x - R_j| > \lambda D_j. \end{aligned}$$

The point about this inequality is that the potential W has the same singularity near each nucleus as V_C , and that the rightmost term in (12) is repulsive. This term will be responsible for stabilizing the system.

The problem is thus reduced to obtaining a lower bound on $q \sum \varepsilon_j$, where $\sum \varepsilon_j$ is the sum of the first $[N/q]$ negative eigenvalues of $\mathcal{T} - W$. Note that Theorem 2 cannot be applied directly to this problem, since W is neither integrable to the power $5/2$ nor to the power 4. Instead we have to do the calculations directly.

For $\nu > 0$ (a number that is chosen later) set $W_\nu(x) \equiv [W(x) - \nu]_+$ and note that $W(x) - \nu \leq W_\nu(x)$. Then, as in (7), $q \sum \varepsilon_j \geq -N\nu - q \int_0^\mu N_{-e}(\mathcal{T} - W_\nu) de$. Again,

$$\int_0^\infty N_{-e}(\mathcal{T} - W_\nu) de \leq \int_0^\mu N_{-e}(\mathcal{T} - W_\nu) de + \int_\mu^\infty N_{-e}(\mu e^{-1} \mathcal{T} - W) de, \quad (13)$$

where we have replaced $W_\nu(x)$ by $W(x)$ in the second term. Applying the CLR bound to the first expression on the right side we obtain $L_3 \int_0^\mu [B(x) + W_\nu(x) - e]_+^{3/2} de d^3x$, which can be bounded, as in part (B), by

$$\sqrt{2} L_3 \int \left\{ \int_0^\mu \left[B(x) - \frac{\gamma e^2}{\mu} \right]_+^{3/2} de + \frac{2}{5} (1 - \gamma)^{-1} W_\nu(x)^{5/2} \right\} d^3x, \quad (14)$$

for any $0 < \gamma < 1$.

The difficulty in dominating the second term in (13) comes from the Coulomb singularity of $W(x)$, which is not fourth power integrable. The singularity can be controlled using the following operator inequality, which follows from the diamagnetic inequality $\int |(\mathbf{p} + \mathbf{A})\psi|^2 d^3x \geq \int |\mathbf{p}\psi|^2 d^3x$ and Lemma 2a on p. 708 of [13].

$$(\mathbf{p} + \mathbf{A})^2 - \frac{Z}{|x|} \geq - \begin{cases} Z^2/4 + \frac{3}{2} Z R^{-1}, & \text{if } |x| \leq R, \\ Z|x|^{-1}, & \text{if } |x| \geq R. \end{cases}$$

Choose $R = \lambda D_j$ and write $(\mathbf{p} + \mathbf{A})^2 = \beta(\mathbf{p} + \mathbf{A})^2 + (1 - \beta)(\mathbf{p} + \mathbf{A})^2$ for some $0 < \beta < 1$. Then, by scaling,

$$(\mu/e)\mathcal{T} - W \geq (\mu/e)(1 - \beta)(\mathbf{p} + \mathbf{A})^2 - (\mu/e)B - \tilde{W},$$

where $\tilde{W}(x, e) = \tilde{G}_j(x, e) + F_j(x)$ for $x \in \Gamma_j$ with $\tilde{G}_j(x, e)$ defined by

$$\begin{aligned} Z^2 e / 4\beta\mu + 3Z / 2\lambda D_j & \text{ for } |x - R_j| \leq \lambda D_j, \\ Z|x - R_j|^{-1} & \text{ for } |x - R_j| > \lambda D_j. \end{aligned}$$

Note that $\tilde{W}$ depends on e .

Again, as in part (B), we can use the CLR bound on the second term in (13) to obtain (when $1 - \gamma \geq Z^2/4\beta\mu$)

$$\begin{aligned} \sqrt{2} L_3 (1 - \beta)^{-3/2} \int \left\{ \int_\mu^\infty [B(x) - \gamma e^2/\mu]_+^{3/2} de \right. \\ \left. + \mu^{-3/2} \int_0^\infty [e \tilde{W}(x, e) - (1 - \gamma)e^2]_+^{3/2} de \right\} d^3x. \end{aligned} \quad (15)$$

First we compute the last integral in (15), which is

$$\begin{aligned} \sum_{j=1}^K \int_{\Gamma_j} \int_0^\infty [e \tilde{G}_j(x, e) + e F_j(x) \\ - (1 - \gamma)e^2]_+^{3/2} de d^3x. \end{aligned}$$

Now split the Γ_j integral into an inner integral $|x - R_j| \leq \lambda D_j$ and an outer integral $|x - R_j| > \lambda D_j$. The inner integral yields, using the definitions of $\tilde{G}_j$ and F_j ,

$$\begin{aligned} \frac{3\pi^2}{32} \left(1 - \gamma - \frac{Z^2}{4\beta\mu} \right)^{-5/2} \int_0^\lambda \left[\frac{1}{2(1 - r^2)} \right. \\ \left. + \frac{3Z}{2\lambda} \right]^4 r^2 dr D_j^{-1}. \end{aligned} \quad (16)$$

To bound the outer integral from above we replace Γ_j by R^3 and get

$$(3\pi^2/32) (1 - \gamma)^{-5/2} (\sqrt{Z} + \sqrt{1/2})^8 (\lambda D_j)^{-1}. \quad (17)$$

Combining (14)–(17) we find that the sum of the negative eigenvalues of $\mathcal{T} - W_\nu$ is bounded below by

$$-a \int W_\nu(x)^{5/2} d^3x - b \int B(x)^2 d^3x - c \sum_{j=1}^K D_j^{-1}. \quad (18)$$

Here $a = q(2\sqrt{2}/5)L_3(1 - \gamma)^{-1}$,

$$b = q \frac{3\pi\sqrt{2}}{16} L_3(1 - \beta)^{-3/2} (\mu/\gamma)^{1/2},$$

$$\begin{aligned} c = q \frac{3\pi^2\sqrt{2}}{32} L_3(1 - \beta)^{-3/2} \mu^{-3/2} \\ \times \left\{ \frac{(\sqrt{Z} + \sqrt{q/4})^8}{\lambda(1 - \gamma)^{5/2}} + \left(1 - \gamma - \frac{Z^2}{4\beta\mu} \right)^{-5/2} \right. \\ \left. \times \int_0^\lambda \left[\frac{q}{4(1 - r^2)} + \frac{3Z}{2\lambda} \right]^4 r^2 dr \right\}. \end{aligned}$$

To simplify the stability condition we have artificially increased the bounds by recalling that $q \geq 2$ and twice replacing $1/2$ by $q/4$ in the definition of c . We choose $\beta = 1/8$, $\gamma = 1/2$, $\lambda = 8/9$, and μ so that $b = (8\pi\alpha^2)^{-1}$. The stability condition $c \leq Z^2/8$ [see (12)] now depends only on the two parameters $X = qZ\alpha^2$ and $Y = q\alpha$. A straightforward but lengthy calculation shows that the stability condition holds if $X = X_0 \equiv 0.082$ and $Y = Y_0 \equiv 0.12$. The condition is monotone in Y , so it holds for $X = X_0, Y \leq Y_0$. Although our condition does not hold for all $X \leq X_0, Y \leq Y_0$, we can use the Z monotonicity of E to conclude stability in this range; this proves (4). With the same values of β , γ , and λ and with $q = 2$ the values $Z = 1050$, $\alpha = 1/137$ also give stability.

To derive (3), note that $W(x) \leq Q|x - R_j|^{-1}$ for $x \in \Gamma_j$. Using this bound and replacing Γ_j by R^3 , one easily obtains $-\sqrt{2}\pi^2 L_3 q K Q^3 \nu^{-1/2} - N\nu$ as a lower bound on the $-a \int W_\nu^{5/2}$ term in (18). Optimizing over ν yields (3) when $X = X_0, Y \leq Y_0$. In this case, $Z \geq Z_0 \equiv 5.7q$. If $X \leq X_0, Y \leq Y_0$ and $Z \geq Z_0$, we get a lower bound on E by increasing α until $X = X_0, Y \leq Y_0$; this yields (3) with $Q = Q(Z)$. Otherwise, with $Z < Z_0$, we use the Z monotonicity of E to conclude (3) with $Q = Q(5.7q)$.

This work was partially supported by NSF Grants No. PHY90-19433-A04 (E. H. L.), No. DMS92-07703 (M. L.), and No. DMS92-03829 (J. P. S.).

- [1] J. Fröhlich, E. Lieb, and M. Loss, *Commun. Math. Phys.* **104**, 251 (1986).
- [2] E. Lieb, *Bull. Am. Math. Soc.* **22**, 1 (1990).
- [3] E. Lieb and H.-T. Yau, *Commun. Math. Phys.* **118**, 177 (1988); *Phys. Rev. Lett.* **61**, 1695 (1988).
- [4] M. Loss and H.-T. Yau, *Commun. Math. Phys.* **104**, 283 (1986).
- [5] J. Avron, I. Herbst, and B. Simon, *Duke Math. J.* **45**, 847 (1978); *Commun. Math. Phys.* **79**, 529 (1981).
- [6] I. Daubechies and E. Lieb, *Commun. Math. Phys.* **90**, 497 (1983).
- [7] E. Lieb and M. Loss, *Commun. Math. Phys.* **104**, 271 (1986).
- [8] J. Conlon, *Commun. Math. Phys.* **94**, 439 (1984).
- [9] C. Fefferman and R. de la Llave, *Rev. Math. Iberoamericana* **2**, 119 (1986).
- [10] J. Combes, R. Schrader, and R. Seiler, *Ann. Phys. (N.Y.)* **111**, 1 (1978).
- [11] E. Lieb, *Proc. Am. Math. Soc. Symposia Pure Math.* **36**, 241 (1980).
- [12] E. Lieb and W. Thirring, in *Studies in Mathematical Physics*, edited by E. H. Lieb, B. Simon, and A. Wightman (Princeton Univ. Press, Princeton, 1976), p. 269.
- [13] A. Lenard and F. Dyson, *J. Math. Phys.* **9**, 698 (1968).

The Chandrasekhar Theory of Stellar Collapse as the Limit of Quantum Mechanics

Elliott H. Lieb* and Horng-Tzer Yau**

Departments of Mathematics and Physics, Princeton University, P.O.B. 708, Princeton, NJ 08544, USA

Dedicated to Walter Thirring on his 60th birthday

Abstract. Starting with a “relativistic” Schrödinger Hamiltonian for neutral gravitating particles, we prove that as the particle number $N \rightarrow \infty$ and the gravitation constant $G \rightarrow 0$ we obtain the well known semiclassical theory for the ground state of stars. For fermions; the correct limit is to fix $GN^{2/3}$ and the Chandrasekhar formula is obtained. For bosons the correct limit is to fix GN and a Hartree type equation is obtained. In the fermion case we also prove that the semiclassical equation has a unique solution – a fact which had not been established previously.

Historical Remarks and Background

There are two principal elementary models of stellar collapse: neutron stars and white dwarfs. In the former there is only one kind of particle which, since it is electrically neutral, interacts only gravitationally. The typical neutron kinetic energy is high, however, so it must be treated relativistically. Unfortunately, the mass and density are also large enough that general relativistic effects are important. For white dwarfs, on the other hand, there are two kinds of nonneutral particles: electrons and nuclei. Because the density is not too large, it is a reasonable approximation to ignore general relativistic effects (although these effects might be important for stability considerations [29]); the nuclei (because of their large mass) can be treated nonrelativistically but the electrons must be treated relativistically. The Coulomb interaction is usually accounted for by the simple assumption that local neutrality requires the nuclear charge density to be equal to the electron charge density, in which case the problem reduces to calculating the electron density. (There are, in fact, electrostatic exchange and correlation effects [28, 29], but these are small by a factor $\alpha = 1/137$.)

* Work partially supported by U.S. National Science Foundation grant PHY 85-15288-A01

** Work supported by Alfred Sloan Foundation dissertation Fellowship

Under the *assumption* of local neutrality (and no significant exchange and electron-nuclei correlation effects) and neglecting the nuclear kinetic energy, the white dwarf problem is mathematically the same as the neutron star problem – but without general relativistic effects. This problem was formulated by Chandrasekhar in 1931 [2] (and also in [7, 11]) and leads to an equation for the density which we here call the Chandrasekhar equation (1.16, 1.18). The neutron star problem leads to the much more complicated Tolman-Oppenheimer-Volkoff equation which will not concern us. Both are reviewed in [24, 27]. Both equations predict collapse at some critical mass which, in the white dwarf case, is called the Chandrasekhar mass. Clearly, near this mass the elementary theory is not totally adequate.

Quantum mechanics is essential for the stability in both cases. “The black-dwarf material is best likened to a single gigantic molecule in its lowest quantum state” [7]. In all treatments up to now, quantum mechanics enters only through the use of a local equation of state $P(\varrho)$, (P = pressure, ϱ = density) which is that of a degenerate Fermi gas (electrons or neutrons). See [30] for example.

Two years ago Lieb and Thirring [19] decided to investigate whether, starting from the Schrödinger equation for *fermions* one would, indeed, recover the semiclassical Chandrasekhar equation (1.16, 1.18) in the limit N (= particle number) $\rightarrow \infty$ and G (= gravitational constant) $\rightarrow 0$. More precisely, for fermions the relevant stability parameter should be $GN^{2/3}$, and not GN . Numerically, the critical N is about 10^{57} , so the limit $N \rightarrow \infty$ is a very reasonable one to consider. The Chandrasekhar value of the critical mass (with the correct $2/3$ exponent) was proved in [19], but only up to a factor of 4. For *bosons*, on the other hand, which have not been considered for astrophysics, Ruffini and Bonazzola [30], Thirring [25], and Messer [21] realized that the relevant parameter should be GN , thus leading to collapse of objects only the size of a mountain. In [19] it was conjectured that, for bosons, (1.18) should be replaced by a Hartree type equation when $N \rightarrow \infty$. In a sense this would mean there is no semiclassical limit for bosons (although we shall continue to employ that word) because the Hartree energy involves density gradients, and not just an equation of state. In [19] the Hartree value of the collapse constant was proved to be correct up to a factor of 2.

In this paper we shall prove that the Chandrasekhar (respectively Hartree) equations are exactly correct as $N \rightarrow \infty$, $G \rightarrow 0$, for *all values* of $GN^{2/3}$ (respectively GN), not just the critical value. In view of Walter Thirring’s contributions to, and interest in quantum mechanical stability questions – in particular the stellar collapse problem – it is a great pleasure for us to dedicate this work to him on the occasion of his 60th birthday.

At first it seemed to us that reducing the quantum problem to a semiclassical problem would end the story. But then we realized that a thorough mathematical study of (1.18), e.g. uniqueness of the solution, has not been done. This, it turns out, is in many ways more complicated than the quantum problem, and therefore a large part of this paper is devoted to an analysis of the semiclassical equations.

In Sect. I we state these problems precisely and summarize the main results. Section II contains proofs of the convergence of the quantum energies to the semiclassical energies. The analysis of the semiclassical equations (existence and uniqueness of solutions and qualitative properties) is in Sect. III and IV. The convergence of the quantum density (for fermions) to the semiclassical density is given in Sect. V.

I. Formulation of the Problem and Main Results

Our starting point is the “relativistic” Schrödinger Hamiltonian for N gravitating particles of mass m (in units $\hbar=c=1$)

$$H_{\kappa N} = \sum_{i=1}^N \{ (p_i^2 + m^2)^{1/2} - m \} - \kappa \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} . \quad (1.1)$$

Here $p^2 = -\Delta$ and $x_i \in \mathbb{R}^3$. $H_{\kappa N}$ can describe a “neutron star” *without* general relativistic effects if we take m = neutron mass and $\kappa = Gm^2$. White dwarfs cannot be described by (1.1) (unless exchange and correlation effects are ignored). A more complicated Hamiltonian is needed in that case and we refer to [19, Sect. 4] for a discussion. Our methods can be extended to the case of several kinds of particles with different masses, but *without* electrostatic interaction. If electrostatic interactions are present, as in white dwarfs, genuinely new ideas are needed. However, if the positive nuclei are *also* fermions, then one can use the inequalities in [19, Sect. 4] to give a lower bound to the energy; unfortunately, this bound will not be the sharp one. It is believed that the semiclassical equation for white dwarfs is nearly the same as for (1.1) as $N \rightarrow \infty$, $G \rightarrow 0$ provided we take $\kappa = G(m + M/z)^2$ with M = nuclear mass, m = electron mass and z = nuclear charge.

For fermions (e.g. neutrons or electrons) $H_{\kappa N}$ acts on antisymmetric functions of space and spin. For generality we assume q spin states/particle; $q=2$ in nature, but $q=1$ would correspond to spin-polarized matter. We also consider $H_{\kappa N}$ without any symmetry restriction. Since the absolute ground state is always symmetric, this is the same as bosons (axion stars?). Technically, $H_{\kappa N}$ is considered as the Friedrich extension of the operator (1.1) with domain $\{ \psi \in L^2(\mathbb{R}^{3N}) | \psi \text{ satisfies fermi statistics (or no statistics in boson case) and } (-\Delta_i)^{1/4} \psi \in L^2(\mathbb{R}^{2N}) \text{ for } i=1 \dots N \}$.

The difficulty in going from $H_{\kappa N}$ to the semiclassical Chandrasekhar or Hartree theories as $N \rightarrow \infty$ and $G \rightarrow 0$ is this: For *one particle*, the operator $h = |p| - Z/|x|$ becomes unbounded below [5, 8–10, 26] when $Z > 2/\pi$. Suppose that, by some fluctuation, $3(\pi\kappa)^{-1}$ particles get very close together. Then they form a trap into which the other particles can fall. Hence we might expect important correlation effects or even collapse for $H_{\kappa N}$ when $N = O(\kappa^{-1})$, in which case the semiclassical point of view wherein the gravitational interaction is treated as a smooth perturbation would be wrong. Something like this *does* happen for bosons and that is why the Hartree equation is the appropriate limiting description. But the interesting (and difficult to prove) fact is that the Pauli principle prevents this from happening for fermions. There is a collapse in that case, but only when $N = O(\kappa^{-3/2})$. The “local equation of state” point of view is valid for fermions.

The quantum energy is defined by

$$E_{\kappa}^Q(N) = \inf \text{spec } H_{\kappa N} \quad (1.2)$$

in the appropriate space according to the statistics. Later on we shall define the quantum density.

The semiclassical energy functionals, $\mathcal{E}$ from $L^1(\mathbb{R}^3)$ to $\mathbb{R}$ are defined as follows.

Fermions: For $t \in \mathbb{R}^+$, let $\eta = (6\pi^2 t/q)^{1/3}$ and

$$\begin{aligned} j(t) &= q(2\pi^2)^{-1} \int_0^{\eta} p^2 \{p^2 + m^2\}^{1/2} - m \} dp \\ &= q(16\pi^2)^{-1} \{ \eta(2\eta^2 + m^2)(\eta^2 + m^2)^{1/2} - m^4 \ln [\eta + (\eta^2 + m^2)^{1/2}] \} - tm . \end{aligned} \quad (1.3)$$

Then

$$\mathcal{E}_{\kappa}^C(\varrho) \equiv \int j(\varrho(x)) dx - \kappa D(\varrho, \varrho) , \quad (1.4)$$

where D is the classical gravitational energy

$$D(\varrho, \varrho) = \frac{1}{2} \iint \varrho(x)\varrho(y)|x-y|^{-1} dx dy . \quad (1.5)$$

$j(\varrho)$ is the ground state kinetic energy density of q -state fermions at density ϱ .

Bosons:

$$\mathcal{E}_{\kappa}^H(\varrho) = (\varrho^{1/2}, \{ (p^2 + m^2)^{1/2} - m \} \varrho^{1/2}) - \kappa D(\varrho, \varrho) . \quad (1.6)$$

The superscript C is for Chandrasekhar while H is for Hartree.

Corresponding to these functionals are the minimum energies:

$$E_{\kappa}^C(N) = \inf \{ \mathcal{E}_{\kappa}^C(\varrho) | \varrho \geq 0, \varrho \in L^{4/3}(\mathbb{R}^3) \text{ and } \int \varrho = N \} , \quad (1.7)$$

$$E_{\kappa}^H(N) = \inf \{ \mathcal{E}_{\kappa}^H(\varrho) | \varrho \geq 0, |p|^{1/2} \varrho^{1/2} \in L^2(\mathbb{R}^3) \text{ and } \int \varrho = N \} . \quad (1.8)$$

Later, we shall omit the subscript κ when it is not necessary.

Recall (see [1, 19] and Lemma 3 for more details) that there is a critical constant $N_f(\kappa)$ which has the properties that $E_{\kappa}^C(N) = -\infty$ iff $N > N_f(\kappa)$. $N_f(\kappa)$ can be calculated explicitly. Define $\gamma \equiv \frac{3}{4} (6\pi^2/q)^{1/3}$ and $\tau_c \equiv \gamma/\sigma_f$, where $\sigma_f = \sup \{ D(\varrho, \varrho) / \int \varrho^{4/3} | \varrho \geq 0, \varrho \in L^{4/3} \text{ and } \int \varrho = 1 \} \approx 1.092$ (see Appendix A). Then

$$N_f(\kappa) = \tau_c^{3/2} \kappa^{-3/2} \approx 4.38 q^{-1/2} \kappa^{-3/2} . \quad (1.9)$$

For bosons, there also exists a critical number $N_b(\kappa)$ which has the properties $E_{\kappa}^H(N) = -\infty$ iff $N > N_b(\kappa)$ (see [19] and Lemma 4). $N_b(\kappa)$ can be related to $\sigma_b \equiv \sup \{ D(\varrho, \varrho) / (\varrho^{1/2}, |p| \varrho^{1/2}) | \varrho \geq 0, |p|^{1/2} \varrho^{1/2} \in L^2 \text{ and } \int \varrho = 1 \}$ by the formula

$$N_b(\kappa) = \sigma_b^{-1} \kappa^{-1} \equiv \omega_c \kappa^{-1} . \quad (1.10)$$

σ_b is known to satisfy $\pi/4 > \sigma_b > 1/2.7$ (Appendix A).

There are scalings

$$E_{\kappa}^C(N) = t^{3/2} E_{t\kappa}^C(t^{-3/2} N) , \quad E_{\kappa}^H(N) = t E_{t\kappa}^H(t^{-1} N) , \quad (1.11)$$

which are easy consequences of the transformation $\varrho(x) \rightarrow \varrho(t^{-1/2}x)$ (respectively $\varrho(x) \rightarrow \varrho(t^{-1/3}x)$). It is convenient to introduce some normalized quantities. For any $\tau > 0$, let

$$\varepsilon_{\tau}^c(\varrho) = \int j(\varrho) - \tau D(\varrho, \varrho) , \quad (1.12)$$

$$e^c(\tau) = \inf \{ \varepsilon_{\tau}^c(\varrho) | \int \varrho = 1 , \varrho \geq 0 \text{ and } \int \varrho^{4/3} < \infty \} . \quad (1.13)$$

It is easy to see that [with $\tilde{\varrho}(x) \equiv \varrho(N^{1/3}x)$]

$$\mathcal{E}_{\kappa}^C(\varrho) = N \varepsilon_{\tau}^c(\tilde{\varrho}) \quad \text{and} \quad E_{\kappa}^C(N) = N e^c(\tau) , \quad (1.14)$$

where $\tau = N^{2/3}\kappa$. Similarly, we have (with $\omega = \kappa N$)

$$\mathcal{E}_\kappa^H(\varrho) = N\mathcal{E}_\omega^H(\tilde{\varrho}) \quad \text{and} \quad E_\kappa^H(N) = N\mathcal{E}^H(\omega), \quad (1.15)$$

where $\mathcal{E}_\omega^H$ and $\mathcal{E}^H(\omega)$ are defined analogously to (1.12, 1.13).

Obviously, if we expect to have a nice limit as $N \rightarrow \infty$ and $G \rightarrow 0$ we should fix the quantities

$$\tau \equiv \kappa N^{2/3} \text{ (fermions) }, \quad \omega \equiv \kappa N \text{ (bosons) }.$$

Numerically, $\kappa \approx 10^{-38}$ for neutrons or nuclei and N is about 10^{57} for a neutron star or white dwarf, so this limit is quite justified physically.

Our main theorems can now be stated.

Theorem 1 (fermions). *Fix $\tau = \kappa N^{2/3}$ and q with $\tau < \tau_c$. Then*

$$\lim_{N \rightarrow \infty} E_\kappa^Q(N)/E_\kappa^C(N) = 1.$$

If $\tau > \tau_c$ then $\lim_{N \rightarrow \infty} E_\kappa^Q(N) = -\infty$.

Theorem 2 (bosons). *Fix $\omega = \kappa N$ with $\omega < \omega_c$. Then*

$$\lim_{N \rightarrow \infty} E_\kappa^Q(N)/E_\kappa^H(N) = 1.$$

If $\omega > \omega_c$ then $\lim_{N \rightarrow \infty} E_\kappa^Q(N) = -\infty$.

Corollary 1. *Let $N_f^Q(\kappa)$ [respectively $N_b^Q(\kappa)$] be the critical particle number for the stability of (1.1) in the fermion (respectively boson) case, i.e. $N^Q(\kappa) = \sup \{N | E_\kappa^Q(N) > -\infty\}$. Then*

$$1 = \lim_{\kappa \rightarrow 0} N_f^Q(\kappa)/N_f(\kappa) = \lim_{\kappa \rightarrow 0} N_b^Q(\kappa)/N_b(\kappa)$$

if q is fixed in the fermion case.

Remarks. (a) In fact, the errors between $E_\kappa^Q(N)$ and $E_\kappa^C(N)$ [respectively $E_\kappa^H(N)$] can be estimated (see Sect. II). The difference between the quantum and semiclassical critical particle numbers can be bounded for large N as follows

$$(1 + 3q^{2/9}N_f(\kappa)^{-2/9})N_f(\kappa) \geq N_f^Q(\kappa) \geq (1 - 20q^{1/9}N_f(\kappa)^{-1/9})N_f(\kappa),$$

$$(1 + 2N_b(\kappa)^{-1})N_b(\kappa) \geq N_b^Q(\kappa) \geq (1 - 10N_b(\kappa)^{-1/3})N_b(\kappa).$$

(b) It was proved in [19] that $\lim N_f^Q(\kappa)/N_f(\kappa)$ is between 1 and 1/4 (roughly). Likewise, $\lim N_b^Q(\kappa)/N_b(\kappa)$ is between 1 and 1/2.

Theorems 1 and 2 show that we can study $H_{\kappa N}$ by means of its semiclassical approximations, (1.4) and (1.6), and therefore it behooves us to study the latter. Auchmuty and Beals [1] showed that there is a minimizing q for (1.4) for each $N < N_f$ and that this q has compact support. They did not prove uniqueness. Later, Lions [20, Theorem II.2] proved that any minimizing sequence of q 's for (1.4) has (after translation $q(x) \rightarrow q(x+y)$) a strongly convergent subsequence in $L^{4/3}(\mathbb{R}^3) \cap L^1(\mathbb{R}^3)$.

The next two theorems summarize some properties of $E_{\kappa}^C(N)$ and $E_{\kappa}^H(N)$ which were not previously known, but which are physically important. For them, we need the notion of symmetric decreasing functions and rearrangements. For the convenience of the reader, we collect some basic definitions and facts about this subject in Appendix A. Since the functions we are interested in are all symmetric decreasing, we shall abuse notation by writing, say, $\varrho(r)$, $\dot{\varrho}(r) \equiv d\varrho/dr$, etc. with $r \equiv |x|$ for a function $\varrho: \mathbb{R}^3 \rightarrow \mathbb{R}$.

Theorem 3 (fermions). (a) *For each $N < N_f$, there exists a symmetric decreasing minimizer $\varrho_N(x)$ for $E_{\kappa}^C(N)$. It satisfies the Euler-Lagrange equation for some Lagrange multiplier μ :*

$$j'(\varrho(x)) = [\eta^2(x) + m^2]^{1/2} - m = \{\kappa|x|^{-1} * \varrho - \mu\}_+ , \quad (1.16)$$

where $\{f(x)\}_+ \equiv \max(f(x), 0)$ and $\eta(x) = (6\pi^2\varrho(x)/q)^{1/3}$.

(b) *Any minimizing ϱ for $E_{\kappa}^C(N)$ is symmetric decreasing after translation and satisfies (1.16) for some μ .*

(c) *There is no minimizing ϱ for $E_{\kappa}^C(N_f)$ even though $E_{\kappa}^C(N_f)$ is finite.*

(d) *$E_{\kappa}^C(N)$ is a strictly concave, monotone decreasing function which is continuous at the end point, N_f , and $E_{\kappa}^C(N_f) = -mN_f$.*

(e) *Let μ_N be the Lagrange multiplier associated to some minimizer ϱ_N for $N < N_f$. Then the right and left derivatives of $E_{\kappa}^C(N)$ satisfy $(dE_{\kappa}^C/dN)_+ \leq \mu_N \leq (dE_{\kappa}^C/dN)_-$.*

(f) *$\mu_N \rightarrow \infty$ as $N \rightarrow N_f$.*

Remarks. (a) The Euler-Lagrange equation (1.16) is in fact equivalent to the Newtonian limit of the Tolman-Oppenheimer-Volkoff equation ((11.1.13) and (11.3.4) of [27], see also [24]). By differentiating (1.16) with respect to r we have

$$\dot{\varrho} j''(\varrho) = \frac{2\pi^2}{q} [\eta^2(r) + m^2]^{-1/2} \eta(r)^{-1} \dot{\varrho}(r) = -\kappa M(r)/r^2 , \quad (1.17)$$

$$M(r) = 4\pi \int_0^r s^2 \varrho(s) ds .$$

Let $P(r) = \frac{q}{6\pi^2} \int_0^{a(r)} k^4 (k^2 + m^2)^{-1/2} dk$ be the pressure ((11.3.43) of [27]). Then (1.17) can be rewritten as an equation of gravitational-hydrostatic equilibrium:

$$-r^2 \dot{P}(r) = \kappa M(r) \varrho(r) . \quad (1.18)$$

Equation (1.18) is the Newtonian limit of the TOV equation. For historical reasons, we call (1.16), and its equivalent (1.18), the Chandrasekhar equation.

(b) The Euler-Lagrange equation for (1.4) is really $j'(\varrho) - \kappa|x|^{-1} * \varrho - \mu = 0$ when $\varrho(x) > 0$ and ≥ 0 when $\varrho(x) = 0$. But, since $j'(0) = 0$, this is equivalent to (1.16).

(c) (1.16) is equivalent to a second order partial differential equation. See (4.7) and Lemma 8.

(d) Theorem 5(b) improves Theorem 3(e).

Theorem 4 (bosons). (a) *For each $N < N_b$, there exists a symmetric decreasing minimizer $\varrho_N(x)$ for $E_{\kappa}^H(N)$. It satisfies the Euler-Lagrange equation with Lagrange*

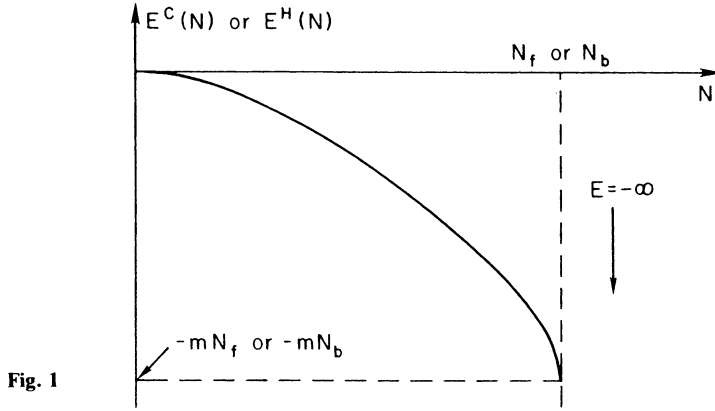


Fig. 1

multiplier v in the distributional sense (with $p^2 = -\Delta$):

$$[(p^2 + m^2)^{1/2} - m]q^{1/2} = (\kappa|x|^{-1} * q - v)q^{1/2}. \quad (1.19)$$

(b), (c), (d), (e), (f) Same as in Theorem 3 mutatis mutandis.

Figure 1, which is schematic, summarizes parts of Theorems 3 and 4.

The fact that $\mathcal{E}_\kappa^C(q)$ has a local kinetic energy enables us to study $E_\kappa^C(N)$ in more detail. In the next theorem, we show that $E_\kappa^C(N)$ has a unique minimizer up to translation for $N < N_f$. We also show that the central density is strictly increasing to infinity while the radius is strictly decreasing to zero as $N \rightarrow N_f$.

The next two theorems are stated only for fermions. While we do not expect that their analogues fail for bosons, to prove them would require a great deal more work. It will be time enough to undertake this work when boson stars are seen in the sky.

Theorem 5 (fermions). (a) For each $N < N_f$ the minimizer q_N is unique up to translations $q(x) \rightarrow q(x+y)$ for $y \in \mathbb{R}^3$.

(b) $E_\kappa^C(N)$ is differentiable in N and thus $dE_\kappa^C(N)/dN = -\mu_N$.

(c) Each q_N has compact support. Let R_N denote the radius of its support. $q_N(r)$ is real analytic for $r < R_N$. R_N is a strictly decreasing function of N . $R_N \rightarrow 0$ as $N \rightarrow N_f$ and $R_N \rightarrow \infty$ as $N \rightarrow 0$.

(d) Let $\alpha_N \equiv q_N(0)$ be the central density of q_N . Then α_N is a strictly increasing function of N tending to ∞ as $N \rightarrow N_f$.

(e) Every radial solution of (1.16) is a minimizer for $N = \int q$.

(f) Any two q_N 's intersect at exactly one value of r .

(g) If $N_1 < N_2$ then $M_1(r) < M_2(r)$ for all $r > 0$ [$M(r)$ is defined in (1.17)].

Some of these results are displayed schematically in Fig. 2.

Remarks. (a) We can relate Theorem 5 to the stability theorem given in e.g. [24, 27]. Since the Euler-Lagrange equation is equivalent to the Chandrasekhar equation, Theorem 5 asserts that any solution of the Chandrasekhar equation with given central density is an absolute (global) minimizer for some $E^C(N)$ or, in other words, is stable. However, our result is stronger than the standard result in [24, 27] where only local stability is discussed.

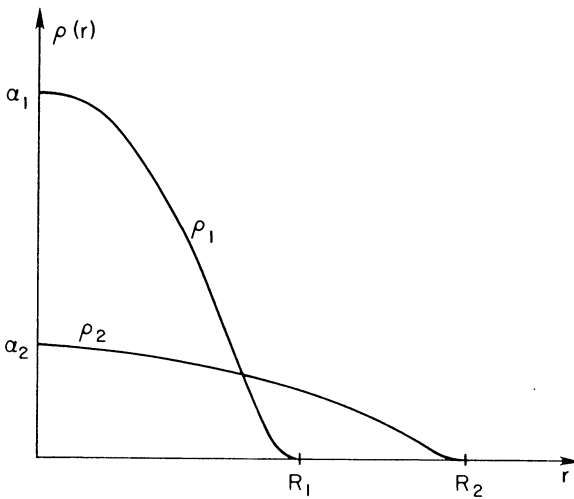


Fig. 2

(b) The normalized semiclassical functional ε_τ^c has properties similar to $\mathcal{E}_\kappa^c$. In particular, $\varepsilon_\tau^c(\tau)$ has a *unique* solution ϱ_τ (say). It is easy to check from the scaling that $\varrho_N(x) = \varrho_\tau(N^{-1/3}x)$.

(c) As will become clear, the uniqueness (up to translations) of the solution amounts to the fact that the radius and central density of the star is a continuous function of the particle number. There are no “phase transitions”. While it is clear that the radius is a continuous function of the central density α , it is not obvious that there are some α ’s which do not correspond to some minimizer. If that were to happen (and we shall show that it does not) then α (and hence the radius) would not be continuous in N .

Equation (1.16) is equivalent to a PDE as shown in (4.7). [Actually, the fact that every solution to (4.7) is a solution to (1.16) is not obvious; in Lemma 8 we prove this for radial solutions.] Theorem 5 translates into statements about the ODE arising from (4.7) in the radial case. This kind of ODE was investigated in an important paper of Ni [23]. He proves that $R(\alpha)$ is a decreasing function of the central density, α , [cf. Theorem 5(c) and (d)] but his uniqueness problem is different from ours since he fixes R while we fix N . It is possible to use Ni’s result to prove Theorem 5 [except for (g)]. We originally followed that route and found the proof to be quite complicated. The proof we actually present here is, in our opinion, much more direct. It uses the variational principle in an essential and, we believe, novel way. Admittedly the class of ODE’s that can be treated by our method is more restricted than Ni’s.

We turn now to the connection between the ground state of $H_{\kappa N}$ and the unique semiclassical density ϱ_N that minimizes $\mathcal{E}_\kappa^c$. Recall that $\varrho_N(x) = \varrho_\tau(N^{-1/3}x)$, where ϱ_τ minimizes ε_τ^c and is independent of N , and $\tau = \kappa N^{2/3}$. Given a normalized N particle function, ψ , of space-spin we define the one-particle density as

$$\varrho_\psi(x) = N \sum_\sigma \int |\psi(x, x_2, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 dx_2 \dots dx_N. \quad (1.20)$$

(A similar definition holds for bosons without $\sum_{\sigma}$.) In some sense, if ψ_N is a “ground state” of $H_{\kappa N}$, and if ϱ_N^Q is the density defined by (1.20) with ψ_N , we expect that $\varrho_N^Q(N^{1/3}x)$ should converge to $\varrho_{\tau}(x)$ as $N \rightarrow \infty$. There are several conceptual difficulties with this expectation as we now explain.

For one thing $H_{\kappa N}$ is translation invariant so it has no L^2 eigenfunction. For this reason, and also because it is physically sensible to consider only functions ψ which are “near” the ground state when the particle number is huge, we first have to introduce the concept of an approximate ground state as in [18].

Definition 1. Fix τ . A sequence of normalized wave functions ψ_N is said to be an *approximate ground state* if, as $N \rightarrow \infty$,

$$N^{-1}|(\psi_N, H_{\kappa N}\psi_N) - E_N^Q| \rightarrow 0. \quad (1.21)$$

Even with this definition there is another problem. In quantum mechanics the fact that ψ_N has a low energy does not imply that the system is localized. To see this, let $g(x)$ be any nonnegative function with $\int g = 1$, and define the density matrix Γ by $\Gamma(X, \sigma | X' \sigma') \equiv \int g(y) \psi_N(x_i + y, \sigma) \psi_N(x'_i + y, \sigma') dy$. One finds easily that the energy of Γ , $\text{Tr } \Gamma H_{\kappa N}$, equals $(\psi_N, H_{\kappa N} \psi_N)$, but Γ has a one-particle density given by $\varrho_{\Gamma} = \varrho_N^Q * g$. Thus the single-particle density can be smeared out as much as we please without affecting the energy.

The usual way out of this difficulty is to fix a certain relatively small number of particles and then to discuss the density of the rest. But it is clear that the number that have to be fixed is huge. Thus, this approach is difficult to implement.

What we do instead is to use a localizing potential. We take advantage of the fact that $\varrho_{\tau}(x)$ has support in a ball of some radius d_{τ} . Let χ_{τ} be the characteristic function of this ball and let $\lambda > 0$. If we add $-\lambda \int \chi_{\tau} \varrho$ to $\mathcal{E}_{\tau}^c$, then the minimum energy would be exactly $e^c(\tau) - \lambda$ and the minimizer would be *uniquely* ϱ_{τ} [formerly all $\varrho_{\tau}(x+y)$, $y \in \mathbb{R}$, were minimizers]. Let us do the same to (1.1)

$$H_{\kappa N \lambda} \equiv H_{\kappa N} - \lambda \sum_{i=1}^N \chi_{\tau}(N^{-1/3} x_i). \quad (1.22)$$

Theorem 6. Fix $\lambda > 0$ and $\tau = \kappa N^{2/3}$. Let $\psi_{N \lambda}$ be a sequence of approximate ground states for $H_{\kappa N \lambda}$ as in (1.21) and let $\varrho_{N \lambda}^Q(x)$ be the densities as in (1.20). Then, for all $\lambda > 0$ and $N \rightarrow \infty$,

$$\varrho_{N \lambda}^Q(N^{1/3} x) \rightarrow \varrho_{\tau}(x) \quad (1.23)$$

weakly in $L^{4/3} \cap L^1(\mathbb{R}^3)$.

Remark. The specific choice of χ_{τ} as a localizing potential is arbitrary. Any other potential for which it is possible to prove the uniqueness of the minimum for $\mathcal{E}_{\tau \lambda}^c$ would suffice.

II. Convergence of the Quantum Energy to the Semiclassical Energy

The easy part is the upper bound to $E^Q(N)$ and we shall dispose of that first. Our proof is basically the same as that in [19, pp. 503–508] which uses the variational principle with coherent states for fermions and a product state for bosons. Only the main points need be given here.

A.1. Upper Bound to the Quantum Energy (Fermions)

To imitate the proof of [19, Theorem 2] it is only necessary to verify the analogue of [19, Eqs. (45), (46)] for our kinetic energy, namely for all $k, p \in \mathbb{R}^3$

$$[k^2 + m^2]^{1/2} - m \leq |k + p| + [p^2 + m^2]^{1/2} - m. \quad (2.1)$$

[This follows from the triangle inequality by thinking of (k, m) , $(p, -m)$ as two vectors in $\mathbb{R}^4$ with the Euclidean norm.] Then we deduce [19, Eq. (51)] that for every nonnegative $\bar{q}(x)$ with $\int \bar{q} = N$ and every $\xi > 0$,

$$E^Q(N) \leq \mathcal{E}_\kappa^C(\bar{q}) + (1.68)\kappa \int \bar{q}^{4/3} + 2N\xi^{1/2} + \xi^{-1}\kappa \int \bar{q}^2. \quad (2.2)$$

Choosing $\bar{q}$ to be a minimizer, q , for $\mathcal{E}_\kappa^C$ and optimizing ξ we obtain

$$E^Q(N) \leq E^C(N) + (1.68)\tau N^{1/3} \int f^{4/3} + 3N^{7/9}\tau^{1/3}(\int f^2)^{1/3}. \quad (2.3)$$

[Recall $\tau = \kappa N^{2/3}$ and $q(x) = f(N^{-1/3}x)$.] Since $E^C(N) = Ne^c(\tau)$ and $f \in L^2 \cap L^\infty$ for a minimizer (cf. Sect. IV) when $\tau < \tau_c$, we see that (2.3) can be bounded for all $\tau < \tau_c$ as

$$E^Q(N) \leq E^C(N)(1 - C_1(\tau, q)N^{-2/9}) \quad (2.4)$$

for some function $C_1(\tau, q)$. Theorem 2 of [19] states that $E^Q(N) = -\infty$ when $\tau > \tau_c(1 + C_2N^{-2/9})$.

A.2. Upper Bound to the Quantum Energy (Bosons)

Again, we follow [19, p. 505]. Given $\bar{q}(x) \geq 0$ with $\int \bar{q} = N$, we define the normalized variational function

$$\psi(x_1, \dots, x_N) \equiv N^{-N/2} \prod_{j=1}^N \bar{q}(x_j)^{1/2}. \quad (2.5)$$

Then, adding and subtracting the self interaction, we get

$$(\psi, H_N \psi) = \mathcal{E}^H(\bar{q}) + \frac{1}{2} \kappa N^{-1} D(\bar{q}, \bar{q}). \quad (2.6)$$

Choosing $\bar{q}$ to be a minimizer, q , for $\mathcal{E}^H$ and recalling that $\omega = \kappa N$, $q(x) = f(N^{-1}x)$, we obtain the analogue of (2.4) for all $\omega < \omega_c$:

$$E^Q(N) \leq E^H(N)(1 - C_3(\omega)N^{-1}). \quad (2.7)$$

When $\omega > [N/(N-1)]\omega_c$, $E^Q(N) = -\infty$.

Now we present the lower bound to $E^Q(N)$ which, apart from the analysis of the semiclassical equation, is the main mathematical point of this paper.

B.1. Lower Bound to the Quantum Energy (Fermions)

As in [19, Eq. (4)] we write H_N as a sum of operators, but here the operators will be more complicated than in [19]. Let P be a partition of $\{1, \dots, N\}$ into two disjoint sets π_1 and π_2 of sizes L and M respectively, with $L + M = N$. There are $\binom{N}{L}$ such partitions. $X = \{x_1, \dots, x_N\}$ denotes the N variables in $\mathbb{R}^3$.

$$H_N = \left\{ \frac{N}{L} \binom{N}{L}^{-1} \sum_P h_P \right\} + H^{(2)} . \quad (2.8)$$

The h_P and $H^{(2)}$ are given in terms of three positive parameters e, κ', ε with $\varepsilon < 1$ as

$$h_P = (1 - \varepsilon) \sum_{i \in \pi_1} \{ (p_i^2 + m^2)^{1/2} - m \} - \kappa' e \left[\sum_{i \in \pi_1} \sum_{j \in \pi_2} |x_i - x_j|^{-1} \right. \\ \left. - e \sum_{j < k \in \pi_2} |x_j - x_k|^{-1} - \sum_{i \in \pi_1} \delta_i(X) - e \sum_{j \in \pi_2} \delta_j(X) \right] , \quad (2.9)$$

$$H^{(2)} = \varepsilon \sum_{i=1}^H \{ (p_i^2 + m^2)^{1/2} - m \} - \kappa' e [1 + eM/L] \sum_{i=1}^N \delta_i(X) . \quad (2.10)$$

In order that (2.8) be an identity we require

$$[2eLM - e^2M(M-1)]\kappa' = L(N-1)\kappa . \quad (2.11)$$

The N functions $\delta_i: \mathbb{R}^{3N} \rightarrow \mathbb{R}$ are defined to be

$$\delta_i(X) = \max \{ |x_j - x_i|^{-1} | j \neq i \} . \quad (2.12)$$

According to Corollary B.2 in the Appendix, $H^{(2)} \geq -\varepsilon mN$ if the following condition among the parameters, which we shall assume, is satisfied.

$$\varepsilon \geq (\pi/2) N^{1/3} \kappa' e [1 + eM/L] . \quad (2.13)$$

Concerning the h_P we note that they are all unitarily equivalent, so it suffices to study one of them. Call the first L variables $Z = \{z_1, \dots, z_L\}$ and the last M variables $Y = \{y_1, \dots, y_M\}$. Since there is no kinetic energy in h_P associated with Y , the y_i can be fixed. Furthermore, for $i = 1, \dots, L$

$$\delta_i(X) \geq \delta(z_i | Y) \equiv \max \{ |z_i - y_j|^{-1} | 1 \leq j \leq M \} \quad (2.14)$$

and, for $i = L + j$ with $1 \leq j \leq M$, $\delta_i(X) \geq \delta_j(Y)$ (since $x_{L+j} = y_j$ by definition). Thus, if we define h^Y on q spin-state fermionic functions of L variables by

$$h^Y \equiv (1 - \varepsilon) \sum_{i=1}^L \{ (p_i^2 + m^2)^{1/2} - m \} - \kappa' e \sum_{i=1}^L V^Y(z_i) + U^Y , \quad (2.15)$$

$$U^Y \equiv \kappa' e^2 \left[\sum_{1 \leq i < j \leq M} |y_i - y_j|^{-1} + \sum_{j=1}^M \delta_j(Y) \right] , \quad (2.16)$$

$$V^Y(z) \equiv \sum_{j=1}^M |z - y_j|^{-1} - \delta(z | Y) , \quad (2.17)$$

we have that for all P

$$h_P \geq \inf_Y \{ \inf \text{spec} (h^Y) \} . \quad (2.18)$$

Lemma 1. *If*

$$\frac{1}{2} > \varepsilon \geq 1.7 q^{1/3} (\kappa')^{2/3} L^{1/3} , \quad (2.19)$$

then for all Y , and with $\kappa'' \equiv (1 - 2\varepsilon)^{-1} \kappa'$,

$$h^Y \geq (1 - 2\varepsilon) E_{\kappa''}^C(L) - \varepsilon m L . \quad (2.20)$$

Proof. Let B_j be the ball centered at y_j of radius $r_j = [2\delta_j(Y)]^{-1}$. These are pairwise disjoint. Let μ_j be the uniform normalized δ -measure on ∂B_j , i.e.

$\mu_j(x) = (4\pi r_j^2)^{-1} \delta(|x - y_j| - r_j)$. Let $\mu = \sum_{j=1}^M \mu_j$. Let ψ be any normalized fermionic L

particle function and $\varrho(x)$ its density [by (1.20) with L in place of N].

Since $D(\varrho - e\mu, \varrho - e\mu) \geq 0$, one easily derives [using $2D(\mu_j, \mu_j) = r_j^{-1}$, $2D(\mu_j, \mu_k) = |y_j - y_k|^{-1}$ and $(|x|^{-1} * \mu_j)(x) = |x - y_j|^{-1}$ for $|x - y_j| \geq r_j$]

$$\kappa' D(\varrho, \varrho) \geq \kappa' e \int \varrho(z) V^Y(z) dz - U^Y. \quad (2.21)$$

Hence, h^Y can be bounded by

$$(\psi, h^Y \psi) \geq (1 - 2\varepsilon) K^Q(\psi) - \kappa' D(\varrho, \varrho) + \varepsilon K^Q(\psi), \quad (2.22)$$

$$K^Q(\psi) \equiv \left(\psi, \left[\sum_{i=1}^L (p_i^2 + m^2)^{1/2} - m \right] \psi \right). \quad (2.23)$$

By Lemma B.3 (Appendix) with $g(x) = \xi^{3/4} \exp(-\pi \xi x^2/2)$ and $(g, |p|g) = 2\xi^{1/2}$, we have

$$\begin{aligned} (1 - 2\varepsilon) [K^Q(\psi) + 2\xi^{1/2} L] - \kappa' D(\varrho * g^2, \varrho * g^2) \\ \geq (1 - 2\varepsilon) \mathcal{E}_{\kappa'}^C(\varrho * g^2) \geq (1 - 2\varepsilon) E_{\kappa'}^C(L). \end{aligned} \quad (2.24)$$

The remainder terms which we have to bound below are

$$R = -2(1 - 2\varepsilon) \xi^{1/2} L + \varepsilon K^Q(\psi) - \frac{1}{2} \kappa' \int \varrho(x) w(x - y) \varrho(y) dx dy \quad (2.25)$$

with $w(x) = |x|^{-1} - (g^2 * |x|^{-1} * g^2)(x)$. The integral in (2.25) can be bounded using Young's inequality by $\|w\|_2 \|\varrho\|_{4/3}^2$. Clearly, $\|w\|_2 = C\xi^{-1/4}$ with C a constant; one easily finds $C^2 \leq 32/(3\pi^{1/2})$. Optimizing the first and last terms in (2.25) with respect to ξ (and replacing $1 - 2\varepsilon$ by 1) we get

$$R \geq -(3/2) C^{2/3} (\kappa')^{2/3} L^{1/3} \|\varrho\|_{4/3}^{4/3} + \varepsilon K^Q(\psi). \quad (2.26)$$

But by (B.10), $K^Q(\psi) \geq 1.6 q^{-1/3} \|\varrho\|_{4/3}^{4/3} - mL$ since $(p^2 + m^2)^{1/2} \geq |p|$. Thus, condition (2.19) implies $R \geq -\varepsilon mL$. $\square$

Let us now put our results together to prove Theorem 1. There are five parameters L, M (with $L + M = N$), ε, e, κ' . These must satisfy (2.11), (2.13), (2.19). We set $e = L/M$ and determine κ' from (2.11), whence $\kappa' < \kappa N/L$. Then we take

$$\varepsilon = 1.7 q^{1/3} \kappa^{2/3} N^{2/3} L^{-1/3} = 1.7 q^{1/3} \tau^{2/3} N^{2/9} L^{-1/3} \quad (2.27)$$

so that (2.19) is satisfied (as $N \rightarrow \infty$, $L/N \rightarrow 1$, so $\varepsilon < 1/2$). The right side of (2.13) is less than $\pi \tau N^{2/3}/M$, so (2.13) will be satisfied [with (2.27)] if we choose

$$M = 1.9 q^{-1/3} \tau^{1/3} N^{7/9}. \quad (2.28)$$

Finally, $L = N - M = N(1 - O(N^{-2/9}))$ and $\varepsilon = O(N^{-1/9})$.

Our lower bound (2.8, 2.20) is thus

$$E_{\kappa'}^Q(N) \geq (N/L) (1 - 2\varepsilon) E_{\kappa'}^C(L) - 2\varepsilon mN \quad (2.29)$$

with $(1 - 2\varepsilon)\kappa'' = \kappa' < \kappa N/L$. Since $N \rightarrow E^C(N)$ is concave and $E^C(0) = 0$ (Sect. III), $(N/L)E^C(L) \geq E^C(N)$, and thus

$$E_k^Q(N) \geq N\{e^c(\tau'') - 2\varepsilon m\} \quad (2.30)$$

with $\tau'' = (1 - 2\varepsilon)^{-1} \tau N/L = \tau + O(N^{-1/9})$. This agrees with $Ne^c(\tau)$ to $O(N^{8/9})$ provided $\tau < \tau_c$. $\square$

B.2. Lower Bound to the Quantum Energy (Bosons)

The proof of Theorem 2 closely follows that of Theorem 1 just given, and only the differences will be mentioned below. Everything from (2.8) to (2.18) is the same. Condition (2.19) is not needed and the replacement for Lemma 1 is

Lemma 2. *Let $\kappa'' = (1 - \varepsilon)^{-1} \kappa'$. Acting on $L^2(\mathbb{R}^{3N})$ without statistics, h^Y satisfies (for all Y)*

$$h^Y \geq (1 - \varepsilon)E_k^H(L) . \quad (2.31)$$

The proof is the same as for Lemma 1 up to (2.22), but now we do not split $K^Q(\psi)$ into two pieces. We merely use Lemma B.5 which immediately yields (2.31).

Our only conditions are (2.11) and (2.13). As before, we set $e = L/M$ and determine κ' from (2.11), giving $\kappa' < \kappa N/L$. To satisfy (2.13) we take

$$\varepsilon = \pi N^{4/3} \kappa / M = \pi N^{1/3} \omega / M . \quad (2.32)$$

(Recall $N\kappa = \omega$ for bosons.) We take $M = N^{2/3}$ so that $\varepsilon = cN^{-1/3}$ and $L/N = 1 - N^{-1/3}$. Then [again using $(N/L)E^H(L) \geq E^H(N)$]

$$E_k^Q(N) \geq N\{e^H(\omega'') - \varepsilon m\} \quad (2.33)$$

with $\omega'' = (1 - \varepsilon)^{-1} \omega N/L$. This agrees with $Ne^H(\omega)$ to $O(N^{2/3})$ when $\omega < \omega_c$.

III. Properties of Semiclassical Energies

We begin with some a priori bounds on the components of the energy.

Lemma 3 (fermions). *For any $q \geq 0$ with $\int q = N \leq N_f$ we have the following a priori estimates:*

$$\int j(q) dx + mN \leq \{1 - (N/N_f)^{2/3}\}^{-1} [\mathcal{E}^C(q) + mN] , \quad (3.1)$$

$$0 \leq E^C(N) + mN \leq CN [1 - (N/N_f)^{2/3}]^{1/2} , \quad (3.2)$$

with $C = \frac{3}{2} m \{ \int (\varrho_F)^{4/3} \int (\varrho_F)^{2/3} \}^{1/2} \|\varrho_F\|_1^{-1}$, where ϱ_F is any minimizer for $F(q)$ in (A.4). In particular, (3.2) implies that $E^C(N)$ is left continuous at N_f and $E^C(N_f) = -mN_f$. When $N > N_f$, $E^C(N) = -\infty$.

Proof. Let $j_0(q) = j(q) + mQ$. By definition [and recalling $\gamma = \frac{3}{4} (6\pi^2/g)^{1/3}$]

$$mN + \mathcal{E}^C(q) = (N/N_f)^{2/3} [\int j_0(q) - (N_f/N)^{2/3} \kappa D(q, q)] + (1 - (N/N_f)^{2/3}) \int j_0(q) .$$

Since $\int \varrho = N$ and $\kappa = \sigma_f^{-1} \gamma N_f^{-2/3}$ we have from the inequality $\int j_0(\varrho) \geq \gamma \int \varrho^{4/3}$ that

$$mN + \mathcal{E}^C(\varrho) \geq (N/N_f)^{2/3} \gamma [\int \varrho^{4/3} - \sigma_f^{-1} N^{-2/3} D(\varrho, \varrho)] + (1 - (N/N_f)^{2/3}) \int j_0(\varrho) \\ \geq (1 - (N/N_f)^{2/3}) \int j_0(\varrho) .$$

The last inequality follows from the definition of σ_f in (A.5) and $\int \varrho = N$.

Let $\varrho_\lambda(x) = \lambda^3 \varrho(\lambda x)$, where $\varrho \geq 0$ is a fixed minimizer for $F(\varrho)$ in (A.4) with $\int \varrho = 1$. By considering $\varrho_\lambda(N^{-1/3} x)$, we have from the variational principle that $mN + E^C(N) \leq N \int j_0(\varrho_\lambda) - N^{5/3} \kappa D(\varrho_\lambda, \varrho_\lambda)$. Since $(p^2 + m^2)^{1/2} \leq |p| + m^2/2|p|$, we find $j_0(\varrho) \leq \gamma \varrho^{4/3} + \frac{9}{16} m^2 \gamma^{-1} \varrho^{2/3}$. Using this,

$$mN + E^C(N) \leq N [\gamma \lambda \int \varrho^{4/3} + 9m^2(16\gamma\lambda)^{-1} \int \varrho^{2/3} - \kappa N^{2/3} \lambda D(\varrho, \varrho)] .$$

By definition, $D(\varrho, \varrho) = \sigma_f \int \varrho^{4/3}$ and $\kappa = \sigma_f^{-1} \gamma N_f^{-2/3}$, whence

$$mN + E^C(N) \leq N \{ [1 - (N/N_f)^{2/3}] \lambda \gamma \int \varrho^{4/3} + 9m^2(16\gamma\lambda)^{-1} \int \varrho^{2/3} \} .$$

Optimizing this with respect to λ yields (3.2) and it also yields $E^C(N) = -\infty$ when $N > N_f$. $\square$

Lemma 4 (bosons). *For any $\varrho \geq 0$ and $\int \varrho = N \leq N_b$, we have the following a priori estimates*

$$(\varrho^{1/2}, (p^2 + m^2)^{1/2} \varrho^{1/2}) \leq (1 - N/N_b)^{-1} (\mathcal{E}^H(\varrho) + mN) , \quad (3.3)$$

$$0 \leq E^H(N) + mN \leq CN [1 - N/N_b]^{1/3} , \quad (3.4)$$

where $C = 2m(\varrho_B^{1/2}, |p|\varrho_B^{1/2})^{1/3} (\varrho_B^{1/2}, |p|^{-1/2} \varrho_B^{1/2})^{2/3} \|\varrho_B\|_1^{-1}$ and $\varrho_B^{1/2}$ is any minimizer for $B(\psi)$ in (A.6). Corollary A.2 implies that $C < \infty$. In particular, $E^H(N)$ is left continuous at N_b and $E^H(N_b) = -mN_b$. When $N > N_b$, $E^H(N) = -\infty$.

Proof. Similar to Lemma 3, one uses $(p^2 + m^2)^{1/2} \leq |p| + m^{3/2}|p|^{-1/2}$. $\square$

In order to prove the concavity of $E^C(N)$ and $E^H(N)$, a technical lemma, whose proof is elementary, is needed.

Lemma 5. *Suppose $f: [a, b] \rightarrow \mathbb{R}$ is a strictly decreasing concave function. Then for any constant $s > 0$ the functions $g(t) = tf(st^{2/3})$ and $h(t) = tf(st)$ are strictly concave on $[(a/s)^{3/2}, (b/s)^{3/2}]$ and $[a/s, b/s]$ respectively.*

Now we can begin the proofs of Theorems 3 and 4.

Proof of Theorems 3(d) and 4(d). $e^c(\tau)$ in (1.13) is a strictly decreasing concave function because $e^c(\tau) = \inf \{ \text{non-constant linear functions of } \tau \}$. The strict concavity of $E^C(N)$ and $E^H(N)$ follows from the scaling relations (1.14, 1.15) and Lemma 5. The continuity of $E^C(N)$ and $E^H(N)$ at N_f and N_b is in Lemmas 3 and 4. $\square$

Proof of Theorems 3(a), (b), (c), (e). The existence of a minimizer was first proved by Auchmuty and Beals [1]. The fact that a minimizing ϱ is necessarily symmetric decreasing follows from the strong rearrangement inequality in [13] (see Appendix A.1).

Part (b) is a standard result in the calculus of variations.

To prove (c) we use the a-priori estimate (3.2) which reads (under the assumption of a minimizing q when $N = N_f$) $-mN_f = E^C(N_f) = \mathcal{E}^C(q)$. Since $(p^2 + m^2)^{1/2} - m > |p| - m$, we have generally (with $\int q = N$)

$$j(q) > \gamma q^{4/3} - mN. \quad (3.5)$$

Thus, in our case, $-mN_f > A(q) - mN_f$ with $A(q) = \gamma \int q^{4/3} - \kappa D(q, q)$. But when $N = N_f = \tau_c^{3/2} \kappa^{-3/2}$, $A(q) \geq 0$ [see (1.9) and (A.5)]. This is a contradiction.

As for (e), the concavity of $E^C(N)$ implies the existence of left-hand and right-hand derivations for all N , and they are equal a.e. Part (e) follows by considering tq_N as a variational function, differentiating $\mathcal{E}^C(tq_N)$ at $t = 1 +$ and $1 -$, and using the Euler-Lagrange equation (1.16). $\square$

Proof of Theorem 4(a), (b), (c), (e). The proof is essentially the same as that for Theorem 3 except for the existence part (a). Here the proof is virtually the same as that for Choquard's equation in [13]. Note that Lemma 4 places an a-priori bound on the kinetic energy which permits us to use a weak compactness argument. The key fact is in (A.2) [and also (A.1)] which permits us to restrict a minimizing sequence to symmetric decreasing functions. Note that the weak compactness argument leads to the existence of a function $\bar{q}_N$ satisfying $\int \bar{q}_N \leq N$ and $\mathcal{E}^H(\bar{q}_N) \leq E^H(N)$. Since $E^H(N)$ is strictly monotone decreasing in N , we must have equality in both cases. $\square$

To prove Theorems 3(f) and 4(f), Lemmas 6 and 7 (which are intrinsically different in the two cases) are needed.

Lemma 6 (fermions). *Let q and $\mu > 0$ satisfy (1.16) with $\int q = N$. We do not assume q is a minimizer for N , but we do assume q is radial, i.e. $q(x) = q(r)$, $r = |x|$. Then*

$$\mathcal{E}^C(q) = \int \{4j(q) - 3qj'(q)\} dx, \quad (3.6)$$

$$K(q) \equiv \int j(q) dx = \frac{5}{2} \mathcal{E}^C(q) + \frac{3}{2} \mu N. \quad (3.7)$$

Proof. Multiply (1.17) by $r^3 q(r)$ and integrate. Then $-4\pi \int_0^\infty r^3 j''(q) q \dot{q} dr = 4\pi \kappa \int_0^\infty M(r) q(r) r dr$. But the second integrals is $\kappa D(q, q)$ (by Newton's theorem).

The first integrals is $-4\pi \int r^3 \frac{d}{dr} (qj'(q) - j(q)) dr$. After integrating by parts [and using $(qj' - j)(0) = 0$] it becomes $3 \int \{qj'(q) - j(q)\} dx$. This proves (3.6) since $\mathcal{E}^C(q) = K(q) - \kappa D(q, q)$. To prove (3.7), multiply (1.16) by $q(x)$ and integrate. Then $\int qj'(q) dx = 2\kappa D(q, q) - \mu N$. Combining this with (3.6) yields (3.7). $\square$

Remark. (3.6) is a *virial theorem*. It can also be proved for minimizing q 's by replacing $q(x)$ by $q_\lambda(x) = \lambda^3 q(\lambda x)$ and differentiating $\mathcal{E}^C(q_\lambda)$ with respect to λ at $\lambda = 1$.

Lemma 7 (bosons). *Suppose $q(x) \geq 0$, $q^{1/2} \in L^2(\mathbb{R}^3)$, $(q^{1/2}, |p|q^{1/2}) < \infty$ and q satisfies (1.19) for some v (in the sense of distributions). Let $\int q = N$. Then*

$$K(q) \equiv (q^{1/2}, \{(p^2 + m^2)^{1/2} - m\} q^{1/2}) = 2\mathcal{E}^H(q) + vN. \quad (3.8)$$

Proof. Since $(\varrho^{1/2}, |p|\varrho^{1/2}) < \infty$ and $\varrho^{1/2} \in L^2$, we have $\int \varrho(|x|^{-1} * \varrho) < (\text{const}) \|\varrho\|_{6/5}^2 < \infty$ by Young's and Sobolev's inequalities. Multiply (1.19) by $\varrho^{1/2}$ and integrate. This yields (3.8). Note that although (1.19) holds only in $\mathscr{D}'$, this integration is justified since all the terms are separately finite. $\square$

Proof of Theorems 3(f), 4(f). We prove the boson case using Lemma 7. The fermion proof is the same using Lemma 6. Assume, on the contrary, that there is a sequence $N \rightarrow N_b$ with minimizers ϱ_N satisfying (1.19) with v_N . Suppose that $v_N \not\rightarrow \infty$. Then since $\mathscr{E}^H(\varrho_N) = E^H(N) \rightarrow E^H(N_b) = -mN_b$ [Theorem 4(d)], we see from Lemma 7 that $K(\varrho_N)$ is bounded. Recall in the proof of Theorem 4(a) that the proof of the existence of a minimizer for any N needed a bound on $K(\varrho)$ for a minimizing sequence. Formerly we used Lemma 4 to achieve this when $N < N_b$. But now, by our assumption on v_N , we also have uniform boundedness of $K(\varrho_N)$. By the proof of Theorem 4(a) we have a function $\bar{\varrho}$ (weak limit of ϱ_N) with $\int \bar{\varrho} \leq N_b$ and $\mathscr{E}^H(\bar{\varrho}) = -mN_b$. As in Theorem 4(a), this implies that $\bar{\varrho}$ is a minimizer for $N = N_b$ and this contradicts Theorem 4(c). $\square$

IV. Properties of the Semiclassical Density (Fermions)

Our main goal here is to prove the uniqueness of the minimizer of the semiclassical functional for each $N < N_f$. This will enable us to prove Theorem 5. The main facts about the density ϱ_N are summarized in Fig. 2, which will be explained later. We shall explore all *radial* [i.e. $\varrho(x) = \varrho(|x|)$] solutions to (1.16); this class includes all minimizers by Theorem 3. Henceforth ϱ will be assumed to be radial without further mention. We shall also suppress irrelevant constants [by replacing $\varrho(x)$ by $a\varrho(bx)$] so that (1.16) becomes

$$j'(\varrho(x)) = [(|x|^{-1} * \varrho)(x) - \mu]_+ \quad (4.1)$$

with $j'(t) = (t^{2/3} + 1)^{1/2} - 1$. The side condition is $\int \varrho = N < N_f = (3/4\sigma_f)^{3/2}$ in these units [see (A.5) and Lemma 3]. We also have defined

$$M(r) = 4\pi \int_0^r \varrho(s) s^2 ds \quad (4.2)$$

so that

$$V_\varrho \equiv |x|^{-1} * \varrho = r^{-1} M(r) + 4\pi \int_r^\infty t \varrho(t) dt. \quad (4.3)$$

Equation (4.1) implies

$$\dot{\varrho} j''(\varrho) = \frac{1}{3} (\varrho^{4/3} + \varrho^{2/3})^{-1/2} \dot{\varrho} = -r^{-2} M(r). \quad (4.4)$$

A. Regularity Properties

Since $V_\varrho(r) \approx N/r$ as $r \rightarrow \infty$ we see, first of all, that $\mu > 0$, for otherwise for large r , $\varrho \approx (N/r)^{3/2} \notin L^1$. By Newton's theorem, (4.1) implies that ϱ has compact support in a ball of radius

$$R = N/\mu \quad (4.5)$$

and, since $V_\varrho(r)$ is continuous in r and $j'(\varrho)$ is continuous in ϱ with $j'(0)=0$, we have that $\varrho(r) \rightarrow 0$ as $r \rightarrow R$. Since $dV_\varrho/dr < 0$, in the domain $B_R = \{r | r < R\}$ we can write (4.1) as

$$j'(\varrho(x)) = V_\varrho(x) - \mu, \quad \text{in } B_R \quad (4.6)$$

with $\varrho(R)=0$. We also see [since $j'(\varrho)$ is monotone] that ϱ is *necessarily* monotone decreasing. Since $V_\varrho(r) < N/r$ we can easily iterate (4.6), starting with $j'(\varrho) < N/r - \mu$, to conclude that ϱ is C^∞ in B_R . (Here we have to use the fact that $(j')^{-1}(t) = (t^2 + 2t)^{3/2}$ is C^∞ for $t > 0$).

By applying Δ to (4.6) we get

$$-\Delta\Theta = 4\pi(j')^{-1}(\Theta) = 4\pi[\Theta^2 + 2\Theta]^{3/2} \quad (4.7)$$

with $\Theta = V_\varrho - \mu \geq 0$. Equation (4.7) holds on B_R with the boundary condition $\Theta(R)=0$. Notice that in this version of the problem N is not mentioned. By [22, Theorem 5.8.6] we can go one step further and assert that Θ (and hence ϱ) is real analytic in B_R .

Since ϱ is radial, (4.7) is an ODE. We are looking for a strong solution to (4.7), with $\Theta(0)=\beta < \infty$ and $\dot{\Theta}(0)=0$. Starting with these initial conditions (with β arbitrary) we can easily (using Picard's method for example) prove that (4.7) has a unique positive solution for each β up to some $R=R(\beta)$ at which $\Theta(R)$ vanishes (which conceivably could be $R=\infty$).

It is clear that every strong solution to (4.7) with the property that $R(\beta) < \infty$ gives a solution to (4.6). Conversely, a solution to (4.6) gives a solution to (4.7) with $\beta=j'(\varrho(0))$. However, it is conceivable that for some $\beta > 0$, the solution to (4.7) has $R(\beta)=\infty$. Later on, by an indirect argument, we will show that this does not happen, but it seems worthwhile to give a direct proof now.

Lemma 8. *Let Θ be the radial solution to (4.7) obtained by integrating outwards from $r=0$ with the initial conditions $\Theta(0)=\beta$ and $\dot{\Theta}(0)=0$. Then there is some $0 < R(\beta) < \infty$ such that $\Theta(R(\beta))=0$. For $r < R(\beta)$, $\dot{\Theta}(r) < 0$.*

Proof. Integrating (4.7) twice we get, for $r \leq R(\beta)$,

$$\Theta(r) = \beta - 4\pi \int_0^r (t^{-1} - r^{-1}) t^2 \varrho(t) dt \quad (4.8)$$

with $\varrho = (\Theta^2 + 2\Theta)^{3/2}$. From (4.8) it is clear that $\dot{\Theta}(r) < 0$. Suppose $R(\beta) = \infty$. We first claim $g(t) \equiv t\varrho(t) \in L^1([0, \infty), dr)$. This follows from $\int_0^r (t^{-1} - r^{-1}) t g(t) dt \geq \int_0^{r/2} (t^{-1} - r^{-1}) t g(t) dt \geq \frac{1}{2} \int_0^{r/2} g(t) dt$; thus, if $g(t) \notin L^1$ we would have from (4.8) that $\Theta(r) < 0$ for large r . Next, by Proposition 9 below, $r^{-1} \int_0^r t g(t) dt \rightarrow 0$ as $r \rightarrow \infty$. Then, from (4.8), as $r \rightarrow \infty$, $\Theta(r) \rightarrow \beta - C$ with $C = 4\pi \int_0^\infty g(t) dt$. Since $\Theta(r) \geq 0$, $\beta \geq C$. Hence $\Theta(r) \geq 4\pi r^{-1} \int_0^r t g(t) dt \geq C'/r$ for large r . But, since $\varrho \geq (2\Theta)^{3/2}$, we have that $\varrho > C'' r^{-3/2}$ for large r and hence $\int_0^\infty g(t) dt = \infty$, which is a contradiction. $\square$

Proposition 9. Suppose $f: [0, \infty) \rightarrow \mathbb{R}^+$ and $\int_0^\infty f(t)dt = C < \infty$. Then $r^{-1} \int_0^r tf(t)dt \rightarrow 0$ as $r \rightarrow \infty$.

Proof. Let $I(r) = \int_0^r tf(t)dt$. Choose $1 > \varepsilon > 0$ and write $I(r) = \int_0^{r\varepsilon} tf(t)dt + \int_{r\varepsilon}^r tf(t)dt$. The first integral is bounded by $r\varepsilon \int_0^{r\varepsilon} f(t)dt \leq r\varepsilon C$. The second is bounded by $rK_\varepsilon(r)$ with $K_\varepsilon(r) = \int_{r\varepsilon}^r f(t)dt$. Since $f \in L^1$, $K_\varepsilon \rightarrow 0$ as $r \rightarrow \infty$. Thus $\limsup r^{-1}I(r) \leq \varepsilon C + \lim_{r \rightarrow \infty} K_\varepsilon(r) = \varepsilon C$. This holds for all ε , and thus proves the Proposition. $\square$

Let us pause to summarize the situation. For each choice of the central density $\varrho(0)$, which we henceforth call α ($\alpha = (\beta^2 + 2\beta)^{3/2}$) there is (by Lemma 8) a unique radial solution that satisfies (4.1) for some unique $\mu = \mu(\alpha) > 0$. This ϱ is real analytic up to $R = N(\alpha)/\mu(\alpha)$, where $N(\alpha) = \int \varrho$. The qualitative nature of this ϱ is shown in Fig. 2.

Suppose that $N(\alpha)$ were a strictly monotone increasing function of α . Then, since a minimizing ϱ satisfies (4.1) and is radial we would conclude: (i) All radial solutions of (4.1) are minimizers; (ii) for each N the minimizing ϱ is unique. But we do not yet know that $N(\alpha)$ is strictly monotone increasing, and that is the problem we now address. Up to this point the arguments were fairly standard (with the possible exception of Lemma 8) and that is why we were brief.

B. Uniqueness and Comparison Properties of Minimizers

Our strategy will be to first focus on solutions to (4.1) which are minimizers for $\mathcal{E}^C$. Then we will show that all solutions to (4.1) are minimizers.

Lemma 10. Suppose ϱ_1 and ϱ_2 are minimizers for $\mathcal{E}^C$ with $\int \varrho_1 = N_1$, $\int \varrho_2 = N_2$ respectively. Let R_1 and R_2 be the radii of their supports and let $R = \max(R_1, R_2)$. Suppose that $\varrho_1(0) > \varrho_2(0)$. [If $\varrho_1(0) = \varrho_2(0)$ then $\varrho_1 \equiv \varrho_2$ and this is uninteresting.] Then for all $0 < r < R$ we have $M_1(r) > M_2(r)$ [see (4.2)].

Proof. Suppose on the contrary, there is an r_0 such that $M_1(r_0) \leq M_2(r_0)$. Since $M_1(r) > M_2(r)$ for r sufficiently small, there exists (by continuity) an $\tilde{r}$, $0 < \tilde{r} \leq r_0$, such that $M_1(\tilde{r}) = M_2(\tilde{r}) \equiv Q$ (say). We first note that $\varrho_1(\tilde{r}) \neq \varrho_2(\tilde{r})$. Otherwise from (4.4) we would have $\dot{\varrho}_1(\tilde{r}) = \dot{\varrho}_2(\tilde{r})$ which together with $\varrho_1(\tilde{r}) = \varrho_2(\tilde{r})$ would imply $\Theta_1(\tilde{r}) = \Theta_2(\tilde{r})$ and $\dot{\Theta}_1(\tilde{r}) = \dot{\Theta}_2(\tilde{r})$. By the uniqueness theorem of ODE, $\Theta_1 \equiv \Theta_2$ which is a contradiction.

Now define the set (for $N_j \geq Q \geq 0$, $j=1,2$) $\mathcal{A}_Q(N_j) = \left\{ \varrho \geq 0 \mid \int_{|x| \leq \tilde{r}} \varrho = Q \text{ and } \int_{|x| \geq \tilde{r}} \varrho = N_j - Q \right\}$ which is subset of $\mathcal{A}(N_j) = \{ \varrho \geq 0 \mid \int \varrho = N_j \}$. From the variational principle, $E^C(N_1) \leq \inf_{\varrho \in \mathcal{A}_Q(N_1)} \mathcal{E}^C(\varrho) \equiv E_Q(N_1)$. But since $\varrho_1 \in \mathcal{A}_Q(N_1)$ by assumption, we have in fact $E^C(N_1) = E_Q(N_1)$. Similarly $E^C(N_2) = E_Q(N_2)$. Now define the following sets (with i, o denoting inside and outside)

$$\begin{aligned} \mathcal{A}^i &= \{ \varrho^i \geq 0 \mid \varrho^i(x) = 0 \text{ if } |x| > \tilde{r} \text{ and } \int \varrho^i = Q \} , \\ \mathcal{A}^o &= \{ \varrho^o \geq 0 \mid \varrho^o(x) = 0 \text{ if } |x| \leq \tilde{r} \text{ and } \int \varrho^o = N_j - Q \} . \end{aligned}$$

It is easy to check that $S: \mathcal{A}^i \times \mathcal{A}_j^o \rightarrow \mathcal{A}_Q(N_j)$ defined by $S(q^i, q^o) = q^i + q^o$ is a bijection. Define a new functional on $\mathcal{A}_j^o$

$$\mathcal{E}_Q^o(q^o) = \int j(q^o) - D(q^o, q^o) - Q \int |x|^{-1} q^o(x) dx.$$

Let $\mathcal{E}_Q^i(q^i) \equiv \mathcal{E}^C(q^i)$. Then, for any $q \in \mathcal{A}_Q(N_j)$ it is easy to check that [with $S^{-1}(q) = (q^i, q^o)$] $\mathcal{E}^C(q) = \mathcal{E}_Q^i(q^i) + \mathcal{E}_Q^o(q^o)$, where we have used Newton's theorem and the definition of $Q = \int_{|x| \leq \tilde{r}} q$.

Let $E_Q^i = \inf \{ \mathcal{E}_Q^i(q^i) | q^i \in \mathcal{A}^i \}$ and $E_{Q,j}^o = \inf \{ \mathcal{E}_Q^o(q^o) | q^o \in \mathcal{A}_j^o \}$. Then by the variational principle, we have $E_Q(N_j) = E_Q^i + E_{Q,j}^o$. Then, since q_j is a minimizer for $E^C(N_j)$, $\mathcal{E}_Q^i(q_j^i) + \mathcal{E}_Q^o(q_j^o) = \mathcal{E}^C(q_j) = E_Q^i + E_{Q,j}^o$. Since $E_Q^i \leq \mathcal{E}_Q^i(q_j^i)$ and $E_{Q,j}^o \leq \mathcal{E}_Q^o(q_j^o)$ by the variational principle, we obtain $E_Q^i = \mathcal{E}_Q^i(q_j^i)$ and $E_{Q,j}^o = \mathcal{E}_Q^o(q_j^o)$. Now let $\tilde{q} \equiv q_1^i + q_2^o$. It is easy to check that $\tilde{q}$ is also a minimizer for N_2 by using $\mathcal{E}_Q^i(q_2^i) = \mathcal{E}_Q^i(q_1^i)$. But $\tilde{q}$ is not continuous at $\tilde{r}$, which violates the regularity of the minimizer proved above. Another way to reach a contradiction is to note that $\tilde{q} \equiv q_2^i + q_1^o$ is a minimizer for N_1 . One of the two functions, $\tilde{q}$ and $\hat{q}$ must be *increasing* at $\tilde{r}$, and this contradicts the symmetric decreasing property of minimizers. $\square$

Remarks. (a) The same method and conclusion apply to minimizers for some other functionals $\mathcal{E}(q)$ which can be written as $\mathcal{E}(q) = \int j(q(x)) dx - D(q, q)$.

(b) Lemma 10 does not say $M_1(R) > M_2(R)$. In fact, we shall later see that this is true, but we do not yet know it. If we knew in advance that $M_1(R) > M_2(R)$ the proof of the following Lemma 11 would be trivial.

Lemma 11. *There exist at most one minimizing q for $E^C(N)$ when $N < N_f$.*

Proof. Suppose, on the contrary, that we have two minimizers q_1 and q_2 with $q_1(0) > q_2(0)$. Let R_1 and R_2 be the radii of their supports. By Lemma 10, $M_1(r) > M_2(r)$ for all $0 < r < \max(R_1, R_2)$. Then $R_1 \leq R_2$, for otherwise $N = M_2(R_2) < M_1(R_2) < N$. Let $J(z) = 4j(z) - 3j'(z)z$, whence $J(z)$ is concave since $J'(z) = -1 + (z^{2/3} + 1)^{-1/2}$ is decreasing. From (3.6), $0 = \mathcal{E}^C(q_1) - \mathcal{E}^C(q_2) = \int [J(q_1) - J(q_2)] dx = 4\pi \int_0^{R_2} [J(q_1(r)) - J(q_2(r))] r^2 dr \leq 4\pi \int_0^{R_2} r^2 (q_1(r) - q_2(r)) J'(q_2(r)) dr$. The last inequality is a consequence of the concavity of J . Integrating the last integral by parts and using the definition of $M(r)$, we have $0 \leq - \int_0^{R_2} (M_1(r) - M_2(r)) J''(q_2(r)) \dot{q}_2(r) dr$. Since $J''(z) < 0$, $\dot{q}_2(r) \leq 0$ and $M_1(r) > M_2(r)$, this last integral is negative, which is a contradiction. $\square$

Remark. Note that the only property of $j(z)$ used in the above proof is the concavity of $J(z)$. Since the concavity of $J(z)$ is equivalent to the convexity of $z \rightarrow j'(z^3)$, Lemma 11 holds for all functionals with $j'(z^3)$ convex.

Lemma 10 says that if $q_1(0) > q_2(0)$ then $N_1 \geq N_2$. But Lemma 11 says that $N_1 = N_2$ is impossible if q_1 is not identical to q_2 . Therefore we have

Corollary 2. *If q_1 and q_2 are minimizers for N_1 and N_2 respectively and if $q_1(0) > q_2(0)$ then $N_1 > N_2$.*

Lemma 12. Let q^α be the unique bounded nonnegative solution of (4.1) with central density $\alpha = q(0)$. Then q is the unique minimizer for $E^C(N)$ with $N = \int q^\alpha$. In other words, all the solutions of (4.1) parametrized by their central density are in fact minima of $E^C(N)$ for some N .

Proof. Let $G = (0, N_f)$ and let $D = \{\alpha | q^\alpha \text{ is a minimizer for some } N \in G\}$. For each $N \in G$ there is a unique minimizing q_N , and hence a unique central density α_N . We let $\Gamma: G \rightarrow D$ denote this map from $N \in G$ to α_N .

(i) Γ is 1:1 by the aforementioned uniqueness of the ODE (4.7) with given initial condition α .

(ii) Γ^{-1} is continuous on D and D is closed in $\mathbb{R}^+$. To prove this we suppose $\alpha_j \rightarrow \alpha \in \mathbb{R}^+$ monotonically. By Corollary 2, $N_j = \Gamma^{-1}(\alpha_j)$ is monotone and bounded, so N_j has a limit N . It is clear that $N < N_f$ for otherwise $\mu_j \rightarrow \infty$ [Theorem 3(f)] but $V_j(r)$ is uniformly bounded [since $q_j(r) < \alpha_j$ and $\int q_j < N_f$]; these facts would contradict (4.1) when r is small. By continuity of the solution of the ODE with respect to α (this follows e.g. from the contraction mapping principle) $q_j(r) \rightarrow q(r)$ pointwise and $q(r)$ satisfies (4.7) with $q(0) = \alpha$. We have to prove that q is a minimizer for N , which will imply $\alpha \in D$ (whence D is closed) and Γ^{-1} is continuous. By (3.1) q_j is uniformly bounded in $L^{4/3}$ so (by passing to a subsequence) $q_j \rightarrow q$ weakly in $L^{4/3}$. Then $\mathcal{E}^C(q) \leq \lim \mathcal{E}^C(q_j) = \lim E^C(N_j) = E^C(N)$. [The lower semicontinuity, namely $\mathcal{E}^C(q) \leq \lim \mathcal{E}^C(q_j)$, follows as in [13], for example. Clearly $\int j(q_j) \leq \lim \int j(q_j)$ since $q_j \rightarrow q$. On the other hand $D(q_j, q_j) \rightarrow D(q, q)$ since q is symmetric decreasing and $q_j \rightarrow q$ pointwise.] Thus, q is a minimizer for N because $\int q \leq N$ while $E^C(N)$ is strictly decreasing; therefore $\int q = N$.

(iii) Γ^{-1} is 1:1 and continuous from D onto G and also D is closed in $\mathbb{R}^+$. Hence Γ^{-1} is a homeomorphism from D to G . Since G is connected, so is D , and therefore D is a closed interval in $\mathbb{R}^+$. Since the only closed interval in $\mathbb{R}^+$ homeomorphic to $(0, N_f)$ is $\mathbb{R}^+$ itself, we conclude that $D = \mathbb{R}^+$. $\square$

Finally we want to make a qualitative comparison of solutions with different N . First a technical lemma is needed.

Lemma 13. Suppose Θ_1 and Θ_2 are two nonnegative solutions of (4.7) with radii R_1 and R_2 with $R_1 < R_2$. Then there exists an $\tilde{r} < R_1$ such that $\Theta_1(\tilde{r}) \geq \Theta_2(\tilde{r})$.

Proof. This is a standard Sturm comparison argument. Assuming, on the contrary, that $\Theta_2(r) > \Theta_1(r)$ for all $r < R_1$, let $u(r) = \Theta_1(r)/\Theta_2(r)$. Then, from (4.7), u satisfies $\Theta_2 \Delta u + 2 \nabla u \cdot \nabla \Theta_2 + (F(\Theta_1) - F(\Theta_2)) \Theta_2 u = 0$ with $F(t) = 4\pi(t^2 + 2t)^{1/2}(t+2)$. Now $u(R_1) = 0$ and $0 \leq u(r) < 1$ for $r < R_1$. Hence $g \equiv F(\Theta_1) - F(\Theta_2) < 0$. At a maximum, $\Delta u \leq 0$, $\nabla u = 0$, but this contradicts $g < 0$. $\square$

Lemma 14. Let q_1 and q_2 be two solutions to (4.1) with central densities $\alpha_1 > \alpha_2$. Then $R_1 < R_2$ and there is precisely one point r in $(0, R_1)$ at which $q_1(r) = q_2(r)$ (as in Fig. 2).

Proof. We know $M_1(r) > M_2(r)$ when $0 < r \leq R \equiv \max(R_1, R_2)$. Suppose $R_1 \geq R_2$. Then integrate (4.4) from $r < R_2$ to R_2 for both q_1 and q_2 . One has $j'(q_i(r)) - j'(q_i(R_2)) = - \int_r^{R_2} s^{-2} M_i(s) ds$ for $i=1, 2$. Since $j'(q_1(R_2)) \geq 0 = j'(q_2(R_2))$ and $M_1(s) > M_2(s)$ for $0 < s < R_2$, we easily conclude $j'(q_1(r)) > j'(q_2(r))$ and hence

$\varrho_1(r) > \varrho_2(r)$ for $0 \leq r < R_2$. If $R_1 > R_2$, then this contradicts Lemma 13, so suppose $R_1 = R_2 = R$. Then, similar to Lemma 13, define $u = \Theta_2/\Theta_1$. By Gauss's theorem $\Theta_j(R) = N_j R^2$, since $\int \varrho_j = N_j$. Then $u(R) = N_2/N_1 \equiv \delta < 1$. As in Lemma 13 (using $\varrho_1 \geq \varrho_2$), u can have no maximum for $r < R$. Thus $u \leq \delta$. However, $N_2 = \int (\Theta_2^2 + 2\Theta_2)^{3/2} \leq \int (\delta^2 \Theta_1^2 + 2\delta \Theta_1)^{3/2} \leq \delta^{3/2} \int (\Theta_1^2 + 2\Theta_1)^{3/2} = \delta^{3/2} N_1$, so $\delta \geq 1$, which is a contradiction. Thus, $R_1 < R_2$.

Since $\alpha_1 > \alpha_2$ the function $f \equiv \varrho_1 - \varrho_2$ which is C^∞ for $t < R_1$ must have at least one zero in $(0, R_1)$. Suppose there is more than one zero. At each zero we have, by (4.4) that $df/dr < 0$ (since $M_1 > M_2$). But it is easy to see that a C^1 function cannot have a negative derivative at *all* its zeros. $\square$

C. Summary

Let us summarize the results of A. and B., for this is Theorem 5.

(i) All radial solutions of the Euler-Lagrange equation (4.1) are in L^1 and are minimizers.

(ii) They are real analytic up to the cut off radius R . For $r \approx R$, $\Theta(r) \approx R - r$ so $\varrho(r) \approx (R - r)^{3/2}$ as in Fig. 2.

(iii) They are parametrized by the central density α which goes from 0 to ∞ . $\mu(\alpha)$ and $N(\alpha)$ are strictly monotone increasing while $R(\alpha)$ is strictly monotone decreasing. As $\alpha \rightarrow \infty$, $\mu(\alpha) \rightarrow \infty$, $N(\alpha) \rightarrow N_f$, $R(\alpha) \rightarrow 0$. As $\alpha \rightarrow 0$, $N(\alpha) \rightarrow 0$, $R(\alpha) \rightarrow \infty$ [to be proved in (vii) below], and hence $\mu(\alpha) = N(\alpha)/R(\alpha) \rightarrow 0$. N, μ, R are continuous in α .

(iv) Any two solutions always have exactly one intersection as in Fig. 2.

(v) For each r , the mass $M(r)$ is an increasing function of α .

(vi) *Proof of Theorem 5(b)*. In the notation of Theorem 3(e), suppose that $(dE^C/dN)_+ \neq (dE^C/dN)_-$ for some $N < N_f$. In this case it is easy to see that μ_N must be discontinuous at N . However, by (4.5) μ_N is continuous since $N \rightarrow \alpha_N$ is continuous (by Lemma 12) and $\alpha \rightarrow R(\alpha)$ is continuous [by continuity of the solution of (4.7) with respect to the "initial data" at $r=0$].

(vii) *Proof of Theorem 5(c)*. The only fact yet to be proved is that $R_N \rightarrow \infty$ as $N \rightarrow 0$ [or equivalently, $R(\alpha) \rightarrow \infty$ as $\alpha \rightarrow 0$]. Suppose on the contrary, $R(\alpha) < R_0$ for all α . Choose a solution Θ_1 of the Lane-Emden equation

$$\Delta \Theta + 4\pi(2\Theta)^{3/2} = 0 \quad (4.9)$$

with $\Theta_1(R_1) = 0$ for some $R_1 > 2R_0$ and $\Theta_1(0) < 1$. Such a Θ_1 always exists since there is a scaling $\Theta(x) \rightarrow \lambda^4 \Theta(\lambda x)$. Let $\beta \equiv \frac{1}{8} \Theta_1(R_0)$ and let Θ_β be the solution of (4.7) with $\Theta_\beta(0) = \beta$. Then $\Theta_\beta(r) < \frac{1}{8} \Theta_1(r)$ and $(\Theta_\beta^2 + 2\Theta_\beta)^{1/2} (\Theta_\beta + 2) < (2\Theta_1)^{1/2}$ for $0 \leq r \leq R_0$ [since Θ_1 and Θ_β are monotone decreasing and $\Theta_1(R_0) = 8\Theta_\beta(0)$]. But this is impossible as can be seen by the argument given in Lemma 13. $\square$

V. Convergence of the Quantum Density to the Semiclassical Density (Fermions)

Here we prove Theorem 6 for fermions. As explained in Sect. I, we first add a fixed single-particle potential $\lambda \chi_t(N^{-1/3}x)$ to $H_{\kappa N}$. (Recall that χ_t is the characteristic function of the support of ϱ_t .) Following the method in [18] we next add an

additional one parameter single-particle potential $\delta W(N^{-1/3}x)$ and differentiate with respect to δ at $\delta=0$. Two facts have to be established: (i) an extension of Theorem 1 to include the potential $\lambda\chi + \delta W$; (ii) the differentiability of the corresponding semiclassical energy with respect to δ . We shall call (i) Theorem 1' and (ii) Lemma 15.

For any $W \in C_0^\infty(\mathbb{R}^3)$ and $\delta \in \mathbb{R}$, define [with $H_{\kappa N \lambda}$ defined in (1.22)]

$$H_{\kappa N \lambda \delta} \equiv H_{\kappa N \lambda} + \delta \sum_{i=1}^N W(N^{-1/3}x_i) , \quad (5.1)$$

$$\begin{aligned} E_{\lambda \delta}^Q(N) &= \inf \text{spec } H_{\kappa N \lambda \delta} , \\ e_{\tau \lambda \delta}^c(\varrho) &= e_{\tau}^c(\varrho) - \lambda \int \chi_{\tau} \varrho + \delta \int W \varrho , \\ e_{\lambda \delta}^c(r) &= \inf \{ e_{\tau \lambda \delta}^c(\varrho) | \varrho \geq 0, \int \varrho = 1 \} . \end{aligned} \quad (5.2)$$

Theorem 1' (fermions). Fix q, λ and $\tau = \kappa N^{2/3}$ with $\tau < \tau_c$. Then $\lim_{N \rightarrow \infty} N^{-1} E_{\lambda \delta}^Q(N) = e_{\lambda \delta}^c(\tau)$.

Proof. (a) *Upper Bound.* Let $\bar{\varrho} = \varrho * g^2$ with $g(x) = \xi^{3/4} \exp(-\pi \xi x^2/2)$. Then we have generally $G \equiv |\int (\varrho(x) - \bar{\varrho}(x)) W(N^{-1/3}x) dx| \leq \int \varrho \|W * \bar{g}^2 - W\|_\infty$, where $\bar{g}(x) = \xi^{3/4} \exp(-\pi \xi x^2/2)$ with $\xi = N^{2/3} \xi$. By the same method as in the proof of Theorem 1, we only have to show that $N^{-1} [1.68 \kappa \int \varrho^{4/3} + 2N \xi^{1/2} + \xi^{-1} \kappa \int \varrho^2 + G] \rightarrow 0$ as $N \rightarrow \infty$. Note that the minimizing ξ for (2.2) is $\xi = \kappa^{2/3}$, whence $\xi = N^{2/9} \tau^{2/3} \rightarrow \infty$. This implies $N^{-1} G = \|W * \bar{g}^2 - W\|_\infty \rightarrow 0$ as $N \rightarrow \infty$.

(b) *Lower Bound.* We only have to show that [cf. (2.25)] $N^{-1} R' \equiv N^{-1} [R + G] \geq N^{-1} [-2 \xi^{1/2} L + \varepsilon K^Q(\psi) - C \kappa' \xi^{-1/4} \|\varrho\|_{4/3}^2 - G]_- \rightarrow 0$ as $N \rightarrow \infty$ (with $|f|_- = \min(0, f)$). Let $\xi = \max \{L^{-4/3} \|\varrho\|_{4/3}^{8/3} (\kappa')^{4/3}, N^{-1/3}\}$. Then $N^{2/3} \xi \equiv \bar{\xi} \geq N^{1/3}$ whence, as in the proof of part (a), $G/N \rightarrow 0$ as $N \rightarrow \infty$. We take ε and κ' exactly the same as in the proof of Theorem 1. For the first three terms, we can use the same estimate as in Lemma 1 if $\xi = L^{-4/3} \|\varrho\|_{4/3}^{8/3} (\kappa')^{4/3}$. If $\xi = N^{-1/3}$ we have that $\|\varrho\|_{4/3} \leq (\text{const}) N^{1/24}$, and then the result is immediate. $\square$

Lemma 15. Fix λ and $\tau < \tau_c$. Then $e_{\lambda \delta}^c(\tau)$ is differentiable at $\delta=0$ and $de_{\lambda \delta}^c(\tau)/d\delta|_{\delta=0} = \int \varrho_{\tau} W$.

Proof. For each $\delta > 0$ choose a function ϱ_δ satisfying $\delta^{-1} |e_{\tau \lambda \delta}^c(\varrho_\delta) - e_{\lambda \delta}^c(\tau)| \rightarrow 0$ as $\delta \rightarrow 0$. By the variational principle, $\delta \int \varrho_{\tau} W \geq e_{\lambda \delta}^c(\tau) - e_{\lambda}^c(\tau) \geq \delta \int \varrho_\delta W + e_{\lambda \delta}^c(\tau) - e_{\tau \lambda \delta}^c(\varrho_\delta)$. To prove the lemma we only have to show that $\int \varrho_\delta W \rightarrow \int \varrho_{\tau} W$ as $\delta \rightarrow 0$. Clearly, $e_{\lambda \delta}^c(\tau)$ is continuous in δ and therefore, by the assumption about ϱ_δ , we see that ϱ_δ is a minimizing sequence, as $\delta \rightarrow 0$, for $e_{\lambda}^c(\tau) = e^c(\tau) - \lambda$. Since $e_{\tau \lambda}^c(\varrho_\delta) \geq e_{\tau}^c(\varrho_\delta) - \lambda$, we have that ϱ_δ is also a minimizing sequence for $e^c(\tau)$ and $\int \varrho_\delta \chi_{\tau} \rightarrow 1$. We also know that the minimizer, ϱ_{τ} , for $e^c(\tau)$ is unique if we center it at the origin (Theorem 5). By Lion's result [20] there exists a sequence $y_\delta \in \mathbb{R}^3$ such that $\varrho_\delta(x + y_\delta) \rightarrow \varrho_{\tau}(x)$ strongly in $L^{4/3} \cap L^1$. But since $\int \chi_{\tau} \varrho_\delta \rightarrow 1$ we must have $y_\delta \rightarrow 0$. $\square$

Proof of Theorem 6. For any approximate ground state $\psi_{N \lambda}$ we have, from the variational principle, that

$$\delta \int \varrho_{N \lambda}^Q(N^{1/3}x) W(x) dx \geq N^{-1} \{E_{\lambda \delta}^Q(N) - E_{\lambda}^Q(N) + E_{\lambda}^Q(N) - (\psi_{N \lambda}, H_{N \lambda} \psi_{N \lambda})\} .$$

By taking the limit $N \rightarrow \infty$ and then $\delta \rightarrow 0$ (using Theorem 1' and Lemma 15), $\lim_{N \rightarrow \infty} \int \varrho_{N\lambda}^Q(N^{1/3}x)W(x)dx = \int \varrho_\tau(x)W(x)$ for any $W \in C_0^\infty(\mathbb{R}^3)$. Hence $\varrho_{N\lambda}^Q(N^{1/3}x) \rightarrow \varrho_\tau(x)$ in weak $L^{4/3}(\mathbb{R}^3)$. The fact that $\varrho_{N\lambda}^Q(N^{1/3}x) \rightarrow \varrho_\tau(x)$ also in weak $L^1(\mathbb{R}^3)$ follows from Lemma III.4 of [18]. $\square$

Appendix A: Some Definitions and Basic Facts

A.1. Symmetric Decreasing Rearrangements

Given a function $\psi : \mathbb{R}^3 \rightarrow \mathbb{C}$, the *symmetric decreasing rearrangement* ψ^* of ψ satisfies $\psi^* : \mathbb{R}^3 \rightarrow \mathbb{R}^+$, $\psi^*(x)$ depends only on $|x|$ and Lebesgue meas $\{x | \psi^*(x) > a\} = \text{Lebesgue meas } \{x | |\psi(x)| > a\}$ for all $a > 0$. ψ is *symmetric decreasing* if $\psi = \psi^*$. It follows that if $j : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ and $\varrho(x) \geq 0$ then $\int j(\varrho^*)dx = \int j(\varrho)dx$. Also $(\psi^*)^2 = (\psi^2)^*$. A particular case of the Riesz inequality states that

$$\iint \varrho(x)\varrho(y)|x-y|^{-1}dxdy \leq \iint \varrho^*(x)\varrho^*(y)|x-y|^{-1}dxdy, \quad (\text{A.1})$$

and the strict rearrangement inequality in [13] states that (A.1) is a strict inequality unless $\varrho(x) = \varrho^*(x-y)$ for some $y \in \mathbb{R}^3$.

We also have that

$$(\psi, |p|\psi) \geq (\psi^*, |p|\psi^*), \quad (\text{A.2})$$

$$(\psi, T\psi) \geq (\psi^*, T\psi^*), \quad (\text{A.3})$$

with $T = (p^2 + m^2)^{1/2} - m = (-\Delta + m^2)^{1/2} - m$. This follows from the proof in [13] and the fact that the kernels $e^{-t|p|}(x, y)$ and $e^{-tT}(x, y)$ are symmetric decreasing functions of $x - y$, as shown in [4].

A.2. Minimizers for the Gravitational Energy

In the fermion case we are concerned with the ratio (for $\varrho \in L^{4/3} \cap L^1$)

$$F(\varrho) = D(\varrho, \varrho) \|\varrho\|_{4/3}^{-4/3} \|\varrho\|_1^{-2/3} \quad (\text{A.4})$$

which is homogeneous in ϱ and dilation invariant. We set

$$\sigma_f \equiv \sup F(\varrho) = 1.092. \quad (\text{A.5})$$

It is a known fact (proved in [17]) that there is a minimizing $\varrho \equiv \varrho_F$ for $F(\varrho)$. It is unique up to scaling, dilation and translation: $\varrho_F(x) \rightarrow a\varrho_F(bx + y)$. It satisfies Emden's equation. The value 1.092 in (A.5) is numerical.

For bosons we consider the ratio

$$B(\psi) = D(|\psi|^2, |\psi|^2) (\psi, |p|\psi)^{-1} \|\psi\|_2^{-2}, \quad (\text{A.6})$$

$$\sigma_b \equiv \sup \{B(\psi) | \psi \in L^2, (\psi, |p|\psi) < \infty\}. \quad (\text{A.7})$$

By (B.10) with $N=1$, $(\psi, |p|\psi) \geq C \|\psi\|_{8/3}^2$ [actually, there is a Sobolev inequality $(\psi, |p|\psi) \geq C \|\psi\|_3^2$]. This, together with the Hardy-Littlewood-Sobolev inequality, $D(\varrho, \varrho) \leq C \|\varrho\|_{6/5}^2$, shows that $B(\varrho)$ is bounded. In fact, the rearrangement inequalities (A.1), (A.2) permit us to imitate the proof in [17] and show that there

is a minimizing ψ for $B(\psi)$ with $\psi \in L^2$ and $(\psi, |p|\psi) < \infty$. See also the proof of Theorem 4(a). We have no precise numerical value for σ_b . However, Theorem 2 states that $\omega_c = 1/\sigma_b$ and, using the results in [19, pp. 503, 504] we have that $\pi/4 > \sigma_b > 1/2.7$.

Another fact we shall need is

Lemma A.1. *Let ψ be any minimizer for $B(\psi)$. Then $\psi \in L^{3/2+\epsilon}$, all $\epsilon > 0$.*

Proof. Note that $\psi = \psi^*$ and satisfies

$$|p|\psi = \{W(x) - v\}\psi \quad (\text{A.8})$$

with $W = (\text{const}) |x|^{-1} * |\psi|^2$ and $v > 0$. Equation (A.8) is just the Euler-Lagrange equation. Since $|\psi|^2 = |\psi^2|^*$ and $|\psi|^2 \in L^1$, $W(x) \rightarrow 0$ as $|x| \rightarrow \infty$. Thus, writing $f \equiv \{W(x) - v\}\psi$ and using $\psi(x) \geq 0$, we have that f_+ has compact support in a ball of some radius R . Since

$$\psi = |p|^{-1} \{ |p|\psi \} = (2\pi^2)^{-1} |x|^{-2} * f \leq (2\pi^2)^{-1} |x|^{-2} * f_+$$

we see that $\psi(x) < c|x|^{-2}$ for $|x| > R$. $\square$

Corollary A.2. *If ψ is a minimizer for $B(\psi)$ then $(\psi, |p|^{-1/2}\psi) < \infty$.*

Proof. $|p|^{-1/2}$ is convolution with $|x|^{-5/2}$. By the Hardy-Littlewood-Sobolev inequality $\psi \in L^{12/7} \Rightarrow (\psi, |p|^{-1/2}\psi) < \infty$. $\square$

Appendix B: Some Inequalities

B.1. Domination of the Nearest Neighbor Coulomb Potential by $|p|$

Let $Y = \{y_1, \dots, y_N\}$ be N arbitrary, but fixed points in $\mathbb{R}^3$ and let $Z = \{z_1, \dots, z_N\}$ with $z_j \in \mathbb{R}$, $z_j \geq 0$ be given. Define

$$\bar{z} = \|z\|_3 = \left\{ \sum z_j^3 \right\}^{1/3}. \quad (\text{B.1})$$

For $x \in \mathbb{R}^3$, the function

$$\omega_{YZ}(x) \equiv \max_j |x - y_j|^{-1} z_j \quad (\text{B.2})$$

can be called the *nearest neighbor Coulomb potential*. Let

$$Q = \{f | f \in L^2(\mathbb{R}^3), \int |\hat{f}(p)|^2 |p| dp < \infty\}, \quad (\text{B.3})$$

where $\hat{f}(p) = \int f(x) \exp(ip \cdot x) dx$ denotes Fourier transform.

Lemma B.1. *For all Y and Z and $f \in Q$*

$$(2\pi)^{-3} \int |\hat{f}(p)|^2 |p| dp \equiv (f, |p|f) \geq \frac{2}{\pi} (\bar{z})^{-1} (f, \omega_{YZ} f). \quad (\text{B.4})$$

Proof. If $N=1$ the lemma is known [8–10, 26] and we shall reduce (B.4) to that case. Let f^* denote the symmetric decreasing rearrangement of f . By (A.2) $(f, |p|f) \geq (f^*, |p|f^*)$. On the other hand $(f, \omega_{YZ} f) \leq (f^*, \omega_{YZ}^* f^*)$. We claim that

$\omega_{YZ}^*(x) \leq \bar{z}|x|^{-1}$, from which (B.4) follows by applying the $N=1$ lemma to f^* . To prove this it is only necessary to note that for all $b > 0$ (with $\mu = \text{Lebesgue measure}$)

$$\begin{aligned} \mu\{x|\omega_{YZ}(x) > b\} &\leq \sum_{j=1}^N \mu\{x|z_j|x-y_j|^{-1} > b\} \\ &= (4\pi/3)b^{-3} \sum_{j=1}^N z_j^3 = \mu\{x|\bar{z}|x|^{-1} > b\}. \quad \square \end{aligned} \quad (\text{B.5})$$

Corollary B.2. *Let $\psi : \mathbb{R}^{3N} \rightarrow \mathbb{C}$ be an N particle function (without any particular statistics) in Q^N . Let $\delta_i : \mathbb{R}^{3N} \rightarrow \mathbb{R}$ denote the N functions defined in (2.12), i.e. $\delta_i(X) = \max\{|x_i - x_j|^{-1} | j \neq i\}$. Then for each i*

$$(\psi, |p_i|\psi) \geq (2/\pi) (N-1)^{-1/3} (\psi, \delta_i\psi). \quad (\text{B.6})$$

Inequality (B.6) holds without regard to statistics. If, on the other hand, ψ is restricted to be a q -state fermionic function, it is possible to prove that

$$\sum_{i=1}^N (\psi, |p_i|\psi) \geq (\text{const}) q^{-1/3} \sum_{i=1}^N (\psi, \delta_i\psi). \quad (\text{B.7})$$

Since (B.7) is not needed here we defer its proof to a forthcoming paper of ours. Fefferman and de la Llave [6] have proved (B.7) for $q=1$ but their method does not appear to be easily generalizable to $q > 1$.

B.2. Semiclassical Lower Bounds to the Kinetic Energy

The single particle kinetic energy operator is $T = (p^2 + m^2)^{1/2} - m$ with $p^2 = -\Delta$. Let ψ be a normalized wave function for N fermions with q spin states each and let

$$K^Q(\psi) = \left(\psi, \sum_{i=1}^N T_i \psi \right) \quad (\text{B.8})$$

be the total kinetic energy. The semiclassical *approximation* to $K^Q(\psi)$ is

$$K^Q(\psi) \approx K(\varrho_\psi) \equiv \int j(\varrho_\psi(x)) dx \quad (\text{B.9})$$

with ϱ_ψ being the single particle density defined by (1.20) and with $j(t)$ given by (1.3).

It is conjectured that $K(\varrho_\psi)$ is, in fact, a lower bound to $K^Q(\psi)$ but no one has proved this. Daubechies [4], using the method in [12], has found two lower bounds of the right form for fermions, the first of which we use here.

$$\left(\psi, \left\{ \sum_{i=1}^N |p_i| \right\} \psi \right) \geq 1.6 q^{-1/3} \int \varrho_\psi(x)^{4/3} dx, \quad (\text{B.10})$$

$$K^Q(\psi) \geq C \int j(C^{-1} \varrho_\psi(x)) dx \quad (\text{B.11})$$

with $C = 9.6$.

Although (B.10) is important for us, we also need a bound similar to (B.9). This is provided by

Lemma B.3 (fermions). *Let $g \in Q$ [cf. (B.3)] with $\|g\|_2 = 1$. Then, for all normalized ψ ,*

$$K^Q(\psi) \geq \int j((\varrho_\psi * |g|^2)) dx - N(g, |p|g) . \quad (\text{B.12})$$

Proof. This is the same as that given in [15, Eqs. (5.14)–(5.22)]. Introduce the coherent states $g_{pq} \in L^2(\mathbb{R}^3)$ by $g_{pq}(x) = g(x - q) \exp(ip \cdot x)$. Here $p, q \in \mathbb{R}^3$. Let π_{pq} be the projection onto g_{pq} . Then for $f \in L^2(\mathbb{R}^3)$

$$(f, f) = (2\pi)^{-3} \iint (f, \pi_{pq} f) dp dq , \quad (\text{B.13})$$

$$(f, (V * |g|^2) f) = (2\pi)^{-3} \iint V(q) (f, \pi_{pq} f) dp dq , \quad (\text{B.14})$$

$$(f, Tf) \geq (2\pi)^{-3} \iint \tilde{T}(p) (f, \pi_{pq} f) dp dq - \|f\|_2^2 (g, |p|g) . \quad (\text{B.15})$$

Here, $\tilde{T}(p) = (p^2 + m^2)^{1/2} - m$ is a function, not an operator. Equations (B.13) and (B.14) are obvious. Inequality (B.15) is easily proved by writing f as a Fourier integral and then using (2.1) with $k \rightarrow p$, $p \rightarrow r - p$. Thus, if h is the operator $T - (V * |g|^2)(x)$ and γ is a positive semidefinite operator with $\text{Tr} \gamma = \lambda$ we have (by making an eigenvector expansion of γ) that

$$\text{Tr} \gamma h \geq (2\pi)^{-3} \iint dp dq [\tilde{T}(p) - V(q)] M(p, q) - \lambda (g, |p|g) \quad (\text{B.16})$$

with $M(p, q) = \text{Tr} \gamma \pi_{pq} \geq 0$. If, in addition, $0 \leq \gamma \leq qI$ then $0 \leq M(p, q) \leq q$ and hence

$$\begin{aligned} \text{Tr} \gamma h &\geq -(2\pi)^{-3} q \iint dp dq [V(q) - \tilde{T}(p)]_+ - \lambda (g, |p|g) \\ &= -(q/6\pi^2) \int (V^2 + 2mV)^{3/2} V + \int j((q/6\pi^2) (V^2 + 2mV)^{3/2}) - \lambda (g, |p|g) . \end{aligned} \quad (\text{B.17})$$

Recall that $j(t)$ depends on q .

Let $\varrho(x) \equiv \gamma(x, x)$ and take $(V^2 + 2mV)^{3/2} = (6\pi^2/q) \varrho * |g|^2$. Noting that $\text{Tr} \gamma (V * |g|^2) = \int V(\varrho * |g|^2)$, (B.17) becomes

$$\text{Tr} \gamma T \geq \int j(\varrho * |g|^2) - \lambda (g, |p|g) . \quad (\text{B.18})$$

To apply this to our case, let γ be given by the kernel

$$\begin{aligned} \gamma(x, y) &= N \sum_{\sigma} \int \psi(x, x_2, \dots, x_N; \sigma_1, \dots, \sigma_N) \\ &\quad \cdot \bar{\psi}(y, x_2, \dots, x_N, \sigma_1, \dots, \sigma_N) dx_2 \dots dx_N . \end{aligned} \quad (\text{B.19})$$

Then $\gamma \geq 0$, $\lambda = \text{Tr} \gamma = N$. The fact that $\gamma \leq qI$ is standard [16]. Inequality (B.18) becomes (B.12). $\square$

Lemma B.3 can trivially be generalized to operators other than T and to $n \neq 3$ in the following way.

Lemma B.4. *Let $\tilde{T}: \mathbb{R}^n \rightarrow \mathbb{R}^+$ satisfy $\tilde{T}(p) \geq 0$ and, for all $p, q \in \mathbb{R}^n$, $\tilde{T}(p) \leq \tilde{T}(p - q) + \tilde{S}(q)$ for a suitable nonnegative function $\tilde{S}: \mathbb{R}^n \rightarrow \mathbb{R}^+$. Let T and S be the corresponding multiplication operators in momentum space. Let $g: \mathbb{R}^n \rightarrow \mathbb{C}$ with $\|g\|_2 = 1$. Then*

$$(\psi, \sum T_i \psi) \geq \int J((\varrho_\psi * |g|^2)(x)) dx - N(g, Sg) . \quad (\text{B.20})$$

Here, $J: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is defined as follows. Let

$$\begin{aligned}\mu(t) &= q(2\pi)^{-n} \text{ meas } \{p | \tilde{T}(p) \leq t\} \\ \beta(t) &= q(2\pi)^{-n} \int_{\tilde{T}(p) \leq t} \tilde{T}(p) dp .\end{aligned}$$

Then

$$J(t) = \beta(\mu^{-1}(t)) , \quad (\text{B.21})$$

where μ^{-1} is the inverse function. (It is not always uniquely defined everywhere, but $\beta \circ \mu^{-1}$ is.)

On the other hand, for *bosons* (or, more generally, particles without any statistics) (B.9) is not a good approximation to $K^Q(\psi)$. The following is a good approximation and also a bound.

Lemma B.5 (bosons).

$$K^Q(\psi) \geq (Q_{\psi}^{1/2}, TQ_{\psi}^{1/2}) . \quad (\text{B.22})$$

The proof of this is that given by Conlon for Lemma 4.2 in [3]. One only has to verify that e^{-tT} has a positive kernel, but this is proved in [4, Remark 1].

Acknowledgements. The authors are grateful to Christopher King and Michael Loss for helpful discussions about this work.

References

1. Auchmuty, J., Beals, R.: Variational solutions of some nonlinear free boundary problems. Arch. Ration. Mech. Anal. **43**, 255–271 (1971). See also Models of rotating stars. Astrophys. J. **165**, L79–L82 (1971)
2. Chandrasekhar, S.: Phil. Mag. **11**, 592 (1931); Astrophys. J. **74**, 81 (1931); Monthly Notices Roy. Astron. Soc. **91**, 456 (1931); Rev. Mod. Phys. **56**, 137 (1984)
3. Conlon, J.: The ground state energy of a classical gas. Commun. Math. Phys. **94**, 439–458 (1984)
4. Daubechies, I.: An uncertainty principle for fermions with generalized kinetic energy. Commun. Math. Phys. **90**, 511–520 (1983)
5. Daubechies, I., Lieb, E.H.: One-electron relativistic molecules with Coulomb interaction. Commun. Math. Phys. **90**, 497–510 (1983)
6. Fefferman, C., de la Llave, R.: Relativistic stability of matter. I. Rev. Math. Iberoamericana **2**, 119–215 (1986)
7. Fowler, R.H.: Monthly Notices Roy. Astron. Soc. **87**, 114 (1926)
8. Herbst, I.: Spectral theory of the operator $(p^2 + m^2)^{1/2} - ze^2/r$. Commun. Math. Phys. **53**, 285–294 (1977); Errata ibid. **55**, 316 (1977)
9. Kato, T.: Perturbation theory for linear operators. Berlin, Heidelberg, New York: Springer 1966. See Remark 5.12, p. 307
10. Kovalenko, V., Perelmutter, M., Semenov, Ya.: Schrödinger operators with $L_w^{1/2}(\mathbb{R}^l)$ potentials. J. Math. Phys. **22**, 1033–1044 (1981)
11. Landau, L.: Phys. Z. Sowjetunion **1**, 285 (1932)
12. Lieb, E.: The number of bound states of one-body Schrödinger operators and the Weyl problem. Proc. Am. Math. Soc. Symp. Pure Math. **36**, 241–252 (1980). See also: Bounds on the eigenvalues of the Laplace and Schrödinger operators. Bull. Am. Math. Soc. **82**, 751–753 (1976)

13. Lieb, E.: Existence and uniqueness of the minimizing solution of Choquard's nonlinear equation. *Stud. Appl. Math.* **57**, 93–105 (1977)
14. Lieb, E.: Variational principle for many-fermions systems. *Phys. Rev. Lett.* **46**, 457–459 (1981); Errata *ibid.* **47**, 69 (1981)
15. Lieb, E.: Thomas-Fermi and related theories of atoms and molecules. *Rev. Mod. Phys.* **53**, 603–641 (1981); Errata *ibid.* **54**, 311 (1982)
16. Lieb, E.: Density functionals for Coulomb systems. *Int. J. Quant. Chem.* **24**, 243–277 (1983)
17. Lieb, E., Oxford, S.: An improved lower bound on the indirect Coulomb energy. *Int. J. Quant. Chem.* **19**, 427–439 (1981)
18. Lieb, E., Simon, B.: Thomas Fermi theory of atoms, molecules and solids. *Adv. Math.* **23**, 22–116 (1977)
19. Lieb, E., Thirring, W.: Gravitational collapse in quantum mechanics with relativistic kinetic energy. *Ann. Phys. (NY)* **155**, 494–512 (1984)
20. Lions, P.-L.: The concentration compactness principle in the calculus of variations; the locally compact case I. *Ann. Inst. H. Poincaré Anal. non lineaire* **1**, 109–145 (1984)
21. Messer, J.: *Lecture Notes in Physics*, Vol. 147. Berlin, Heidelberg, New York: Springer 1981
22. Morrey, C.B., Jr.: *Multiple integrals in the calculus of variations*, Theorem 5.8.6. Berlin, Heidelberg, New York: Springer 1966
23. Ni, W.M.: Uniqueness of solutions of nonlinear Dirichlet problems. *J. Differ. Equations* **5**, 289–304 (1983)
24. Straumann, S.: *General relativity and relativistic astrophysics*. Berlin, Heidelberg, New York: Springer 1984
25. Thirring, W.: Bosonic black holes. *Phys. Lett.* **B127**, 27 (1983)
26. Weder, R.: Spectral analysis of pseudodifferential operators. *J. Funct. Anal.* **20**, 319–337 (1975)
27. Weinberg, S.: *Gravitation and cosmology*. New York: Wiley 1972
28. Hamada, T., Salpeter, E.: Models for zero temperature stars. *Astrophys. J.* **134**, 683 (1961)
29. Shapiro, S., Teukolsky, S.: *Black holes, white dwarfs and neutron stars*. New York: Wiley 1983
30. Ruffini, R., Bonazzola, S.: Systems of self-gravitating particles in general relativity and the concept of an equation of state. *Phys. Rev.* **187**, 1767–1783 (1969)

Communicated by E.H. Lieb

Received March 2, 1987, in revised form March 27, 1987

One-Electron Relativistic Molecules with Coulomb Interaction

Ingrid Daubechies[†]* and Elliott H. Lieb[†]

Departments of Mathematics and Physics, Princeton University, Princeton, NJ 08544, USA

Abstract. As an approximation to a relativistic one-electron molecule, we study the operator $H = (-\Delta + m^2)^{1/2} - e^2 \sum_{j=1}^K Z_j |x - R_j|^{-1}$ with $Z_j \geq 0$, $e^{-2} = 137.04$. H is bounded below if and only if $e^2 Z_j \leq 2/\pi$, all j . Assuming this condition, the system is unstable when $e^2 \sum Z_j > 2/\pi$ in the sense that $E_0 = \inf \text{spec}(H) \rightarrow -\infty$ as the $R_j \rightarrow 0$, all j . We prove that the nuclear Coulomb repulsion more than restores stability; namely $E_0 + 0.069e^2 \sum_{i < j} Z_i Z_j |R_i - R_j|^{-1} \geq 0$. We also show that E_0 is an increasing function of the internuclear distances $|R_i - R_j|$.

Introduction

The problem of “stability of matter” consists in proving that a system of charged particles (electrons and nuclei), interacting electromagnetically, does not collapse. In the framework of nonrelativistic Schrödinger quantum mechanics, with Coulomb interactions between the particles, a first proof of this was given by F. Dyson and A. Lenard [1]. A shorter proof, leading to a much better lower bound on the binding energy per electron, was later given by E. Lieb and W. Thirring [2]. The strategy these proofs followed was first to consider the nuclei fixed (i.e. with infinite mass); the general case (nuclei with finite mass) then follows easily. With the K nuclei at fixed positions $R_1, \dots, R_K$, the problem then consists in proving that 1) the Hamiltonian describing N electrons and K nuclei (including also the repulsion terms between the electrons, and between the nuclei) is bounded below by a constant independent of the R_j ; 2) the energy per particle, i.e. the ground state energy of the system of N

[†] Work partially supported by U.S. National Science Foundation grant PHY-8116101-A01

* On leave from Vrije Universiteit Brussel, Belgium, and from Interuniversitair Instituut voor Kernwetenschappen, Belgium

electrons and K nuclei, divided by $(N + K)$, is bounded below by some constant.

If electrons were bosons, it is known [3] that matter would not be stable: statement 1) above still holds for bosons, but 2) does not.

As a first approximation to a relativistic approach to the problem of stability of matter, we shall study stability for the Hamiltonian

$$\begin{aligned} \sum_{j=1}^N (-\hbar^2 c^2 \Delta_j + m^2 c^4)^{1/2} - \sum_{j=1}^N \sum_{k=1}^K Z_k e^2 |x_j - R_k|^{-1} + \sum_{\substack{i,j=1 \\ i < j}}^N e^2 |x_i - x_j|^{-1} \\ + \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l e^2 |R_k - R_l|^{-1} = H_{N,K}(m, \underline{Z}, \underline{R}), \end{aligned} \quad (1.1)$$

where m is the electron mass, and where again the K nuclei are fixed at the distinct positions $R_1, \dots, R_K$. We shall use the shorthand notation $\underline{Z}, \underline{R}$ for the sets $\{Z_j\}_{j=1}^K, \{R_j\}_{j=1}^K$.

To simplify our expressions, we rescale the x_j and R_k in units of the Compton wave length of the electron $\hbar/mc$. We also rescale the Hamiltonian by mc^2 . In terms of the variables $\tilde{x}_j = (mc/\hbar)x_j$, and with $\tilde{R}_k = (mc/\hbar)R_k$, $\tilde{H}_{N,K} = (mc^2)^{-1}H_{N,K}$ becomes

$$\begin{aligned} \tilde{H}_{N,K}(\underline{Z}, \underline{R}) = \sum_{j=1}^N (-\tilde{\Delta}_j + 1)^{1/2} - \sum_{j=1}^N \sum_{k=1}^K Z_k \alpha |\tilde{x}_j - \tilde{R}_k|^{-1} \\ + \sum_{\substack{i,j=1 \\ i < j}}^N \alpha |\tilde{x}_i - \tilde{x}_j|^{-1} + \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |\tilde{R}_k - \tilde{R}_l|^{-1}, \end{aligned}$$

where $\alpha = e^2/\hbar c \simeq (137.04)^{-1}$.

It is obvious from the definition (1.1) that $H_{N,K}(m, \underline{Z}, \underline{R})$ is bounded below by $H_{N,K}(0, \underline{Z}, \underline{R})$. This will turn out to be very useful in some of our proofs below. We therefore introduce also a rescaled version of $H_{N,K}(0, \underline{Z}, \underline{R})$, with a different scaling, since the Compton wave length is infinity if $m = 0$. We rescale $H(0, \underline{Z}, \underline{R})$ in units $\hbar c/R_0$, where R_0 is an arbitrary length, and scale x and the R_j by R_0 . This yields

$$\begin{aligned} h_{N,K}(\underline{Z}, \underline{R}) = \sum_{j=1}^N (-\tilde{\Delta}_j)^{1/2} - \sum_{j=1}^N \sum_{k=1}^K Z_k \alpha |\tilde{x}_j - \tilde{R}_k|^{-1} \\ + \sum_{\substack{i,j=1 \\ i < j}}^N \alpha |\tilde{x}_i - \tilde{x}_j|^{-1} + \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |\tilde{R}_k - \tilde{R}_l|^{-1}, \end{aligned}$$

where again the tildes indicate the scaled variables. We shall henceforth always work with these two scaled Hamiltonians, and omit the tildes.

It turns out that, unlike the Schrödinger case, even statement 1) above (i.e. the existence of an R -independent lower bound for the ground state energy of $H_{N,K}$) is not straightforward for $H_{N,K}$ even if we restrict ourselves to the case $N = 1$ (1 electron). First of all, there exists only a limited range of the Z_j for which $H_{N,K}$ is bounded below at all. This can already be seen for the simplest case

$$K = 1, N = 1,$$

$$H_{1,1}(Z) = (-\Delta + 1)^{1/2} - Z\alpha|x|^{-1}. \quad (1.2)$$

This operator was studied independently by R. Weder [4] and I. Herbst [5]. A first important fact (which can be found in Kato [6]) is the existence of a critical charge Z_c for the operator $H_{1,1}$, exactly as for the Klein–Gordon or Dirac theories. To be explicit, let

$$H_0 = (-\Delta + 1)^{1/2}, \quad Z_c = 2/(\alpha\pi) \simeq 87.2.$$

Then a) for $Z \leq Z_c$: $H_{1,1} \geq 0$ as a form on $Q(H_0)$.

b) for $Z > Z_c$: $H_{1,1}$ is unbounded below as a form on $Q(H_0)$.

Here $Q(H_0)$ is the form domain of H_0 , which consists of all the functions f in $L^2(\mathbb{R}^3)$ for which $|p|^{1/2}\hat{f}(p)$ is square integrable, where $\hat{f}$ denotes the Fourier transform of f .

If Z is strictly subcritical (i.e. $Z < Z_c$), more information concerning the spectrum of $H_{1,1}(Z)$ is known [4, 5]. The eigenvalues of $H_{1,1}(Z)$ all lie between 0 and 1; as a matter of fact, they are even separated from 0 by a gap which increases with $Z_c - Z$.

$$\sigma_{\text{disc}}(H_{1,1}) \subset [1 - (Z/Z_c)^2]^{1/2}, 1). \quad (1.3)$$

Since $(1 - \Delta)^{1/2} \leq 1 - \Delta/2$, and since $-\Delta/2 - Ze^2|x|^{-1}$ has infinitely many negative eigenvalues, it follows that $H_{1,1}(Z)$ has infinitely many eigenvalues smaller than 1. It turns out, from an argument we give in Sect. 3, that the lowest eigenvalue is nondegenerate, and that the associated ground state is strictly positive.

On the other hand the essential spectrum always starts at 1, and it consists of only an absolutely continuous spectrum: $\sigma_{\text{ess}}(H_{1,1}) = \sigma_{\text{abscont}}(H_{1,1}) = [1, \infty)$.

The operator $h_{1,1}(Z) = (-\Delta)^{1/2} - Z\alpha|x|^{-1}$ has the same critical Z as $H_{1,1}$: it is bounded below by 0 on $Q(|p|)$ if $Z \leq Z_c$, and unbounded below if $Z > Z_c$. Its spectrum is much simpler however: $\sigma(h_{1,1}) = \sigma_{\text{abscont}}(h_{1,1}) = [0, \infty)$.

One can easily extend the proofs in [4, 5] to show that $H_{N,K}$ is bounded below if and only if $Z_j\alpha \leq 2/\pi$ for all j (we assume that all the R_j are distinct). From now on we shall only consider this case. Even under this condition, however, it is not obvious that an R -independent lower bound for $H_{N,K}$ exists.

Let us consider the case $N = 1$:

$$H_{1,K} = H_{1,K}^0 + \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |R_k - R_l|^{-1}, \quad (1.4)$$

where

$$H_{1,K}^0 = (-\Delta + 1)^{1/2} - \sum_{k=1}^K Z_k \alpha |x - R_k|^{-1} \quad (1.5)$$

is the 1-electron Hamiltonian without nuclear repulsion. Suppose that $\sum_{k=1}^K Z_k \alpha > 2/\pi$ (this is possible when $K \geq 2$). It is a simple consequence of the unboundedness below of $H_{1,1}(Z)$ for $Z\alpha > 2/\pi$ that the ground state energy for $H_{1,K}^0$ will tend to $-\infty$ if the internuclear distances shrink to zero:

$$\lim_{\lambda \rightarrow 0} \inf \text{spec } H_{1,K}^0(Z, \lambda R) = -\infty, \quad \sum_{k=1}^K Z_k \alpha > 2/\pi.$$

Under the same conditions however, the nuclear repulsion term in (1.4) tends to $+\infty$, and this may “cancel” the behavior of $E_0 = \inf_{\text{spec}} H_{1,K}^0$. The existence of an $\underline{R}$ -independent lower bound for $H_{1,K}(\underline{Z}, \underline{R})$ is therefore only possible if the nuclear repulsion is strong enough to overcome the “collapsing tendency” present in $H_{1,K}^0$. The fact that this is the case is the main content of the following theorem:

Theorem 1. *Let $h_0 = (-\Delta)^{1/2}$. Suppose $z_k \leq 2/\pi$ for all k , and $R_k \neq R_l$ for all $k \neq l$. Then $h_0 - \sum_{k=1}^K z_k |x - R_k|^{-1}$, considered as a form on $Q(h_0)$, is bounded below by $-3\pi \sum_{\substack{k,l=1 \\ k < l}}^K z_k z_l |R_k - R_l|^{-1}$.*

Since $(-\Delta + 1)^{1/2} \geq (-\Delta)^{1/2}$, the theorem obviously also holds if we replace h_0 by $H_0 = (-\Delta + 1)^{1/2}$.

In terms of the fine-structure constant α and the nuclear charges $Z_k = z_k/\alpha$, Theorem 1 can be rewritten as

$$E_0 + \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |R_k - R_l|^{-1} \geq (1-b) \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |R_k - R_l|^{-1}, \quad (1.6)$$

where $b = 3\pi\alpha \leq 6.88 \times 10^{-2}$. The fact that b is strictly smaller than 1 implies that the total energy (including the nuclear repulsion) increases to $+\infty$ if any two nuclei coincide. Thus the nuclear repulsion is not only strong enough to prevent collapse, it even pushes the nuclei apart!

For

$$h_{1,K}(\underline{Z}, \underline{R}) = h_{1,K}^0(\underline{Z}, \underline{R}) + \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |R_k - R_l|^{-1}, \quad (1.7)$$

with

$$h_{1,K}^0(\underline{Z}, \underline{R}) = (-\Delta)^{1/2} - \sum_{k=1}^K Z_k \alpha |x - R_k|^{-1}, \quad (1.8)$$

we can say even more. By scaling one sees that $h_{1,K}(\underline{Z}, \underline{R})$ and $\lambda h_{1,K}(\underline{Z}, \lambda \underline{R})$ are unitarily equivalent. Theorem 1 implies therefore that in this case the total energy is minimal if the nuclei are infinitely far apart.

Incidentally, note that the exact numerical value of α plays a decisive role. This can be understood by the scaling behavior of $H_{1,K}$. By simple scaling one sees that $H_{1,K}(\underline{Z}, \underline{R})$ is unitarily equivalent to $\lambda H_{1,K}(\lambda^{-1} \underline{Z}, \lambda \underline{R})$, where we define

$$H_{1,K}(\mu; \underline{Z}, \underline{R}) = (-\Delta + \mu^2)^{1/2} - \sum_{k=1}^K Z_k \alpha |x - R_k|^{-1} + \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |R_k - R_l|^{-1}.$$

(The same is true for $H_{N,K}$.) This means essentially that we can scale the mass away: the existence of any $\underline{R}$ -independent bound for $H_{1,K}(1; \underline{Z}, \underline{R})$ implies that the zero-mass Hamiltonian $H_{1,K}(0; \underline{Z}, \underline{R}) = h_{1,K}(\underline{Z}, \underline{R})$ is bounded below by zero, independently of $\underline{R}$. (Since $H_{1,K} \geq h_{1,K}$, the converse statement is trivial.)

This, in turn, implies that the value of α is decisive. Let us illustrate this for the special case $K = 2$, $Z_1 \alpha = Z_2 \alpha = 2/\pi$. On the one hand, if $H_{1,2}(\underline{Z}, \underline{R})$ is bounded

below by an R -independent constant, then $h_{1,2}(Z, R)$ is bounded below by zero. Thus, for all ψ in $Q[(-\Delta)^{1/2}]$,

$$\langle \psi, h_{1,2}^0(Z, R)\psi \rangle + (4/\pi^2\alpha)|R_1 - R_2|^{-1} \|\psi\|^2 \geq 0$$

($h_{1,2}^0$ is $h_{1,2}$ without the nuclear repulsion; see (1.8)). On the other hand, by a scaling argument,

$$\inf \text{spec} [h_{1,2}^0(Z, R)] = -C|R_1 - R_2|^{-1}$$

for some positive constant C , which is completely independent of α . It is here that the numerical value of α comes into play: if α is such that $4/(\alpha\pi^2) \geq C$, then the required bound will hold (for $K = 2$); if $4/(\alpha\pi^2) < C$, the one-electron-two-nucleus system is definitely unstable in this model. By the argument above, this conclusion will hold as well for $H_{1,K}$, i.e. for any positive mass, as for $h_{1,K}$, which has zero mass.

We prove Theorem 1 in Sect. 2.

In Sect. 3 we show that, for one electron, $E_{1,K}^0 = \inf \text{spec} H_{1,K}^0(Z, R)$, and $e_{1,K}^0 = \inf \text{spec} h_{1,K}^0(Z, R)$ are monotone increasing functions of the internuclear distances:

Theorem 2. *Keep the Z_k fixed, where $Z_k\alpha \leq 2/\pi$ for all k . If for all k, l : $|R_k - R_l| \leq |R'_k - R'_l|$, then $E_{1,K}^0(Z, R) \leq E_{1,K}^0(Z, R')$ and $e_{1,K}^0(Z, R) \leq e_{1,K}^0(Z, R')$.*

Our proof borrows its basic ideas from [7] and [8], where the analogous statement for the Schrödinger operator $-\Delta - \sum_{k=1}^K Z_k e^2 |x - R_k|^{-1}$ was proved (first for dilations, i.e. $R'_k = \lambda R_k$ with $\lambda \geq 1$, in [7]; this proof was extended to the general situation $|R'_k - R'_l| \geq |R_k - R_l|$ in [8]).

If all the Z_k are strictly subcritical, i.e. $Z_k\alpha < 2/\pi$ for all k , one can extend a result of [4, 5] and show that $\sigma_{\text{ess}}(H_{1,K}^0) = [1, \infty)$ and analogously $\sigma_{\text{ess}}(h_{1,K}^0) = [0, \infty)$. For any Z (provided all the Z_k are not zero) one can find a variational function (e.g. a suitable exponential) such that $\langle \psi, H_{1,K}^0\psi \rangle < 1$. This shows that for $Z_k < Z_c$, all k , $E_{1,K}^0$ is an isolated eigenvalue of $H_{1,K}^0$. (Actually, by the same argument as for $H_{1,1}$, one sees that $H_{1,K}^0$ has infinitely many eigenvalues smaller than 1.) For $h_{1,K}^0$, matters are slightly different. If $\sum_{k=1}^K Z_k \leq Z_c$, one can easily extend arguments of [4, 5] to show that $h_{1,K}^0 \geq 0$, in which case $e_{1,K}^0 = 0$, and $h_{1,K}^0$ has no eigenvalues. If $\sum_{k=1}^K Z_k > Z_c$, one can again find a variational function ψ for which $\langle \psi, h_{1,K}^0\psi \rangle < 0$, which implies that in that case $e_{1,K}^0$ is also an isolated eigenvalue.

We prove in Sect. 3 that, if $Z_k < Z_c$ for all k , $E_{1,K}^0$ is a nondegenerate eigenvalue of $H_{1,K}^0$, and that the corresponding ground state is strictly positive. The same is true for $e_{1,K}^0$ under the additional assumption $\sum_{k=1}^K Z_k > Z_c$.

We have no results in this paper concerning $H_{N,K}$ for $N > 1$. The same scaling argument as for $H_{1,K}$ applies, and the existence of an R -independent lower

bound for $H_{N,K}$ would therefore imply that $h_{N,K}$ (i.e. the same Hamiltonian with $(-\Delta)^{1/2}$ instead of $(-\Delta + 1)^{1/2}$) is bounded below by zero. We do not know, however, whether such a lower bound exists for general N, K .

Regardless of the statistics, and neglecting the electron repulsion, we know that if $K = 1$ (one nucleus only) $H_{N,1} \geq 0$ (because $H_{1,1} \geq 0$). On the other hand, the lower bound (1.6) shows that for $N \leq 14$ and arbitrary K , we also have $H_{N,K} \geq 0$ (because $1 - 14b \geq 0$, where $b = 3\pi\alpha$). For general values of N, K it is obvious that the Fermi statistics and the electron repulsion will have to play a role.

It is clear that the strategy of [2], which used a density functional lower bound on the kinetic energy, and then applied results of Thomas–Fermi theory, will not work here. The lower bound on the kinetic energy for the operator $(-\Delta + 1)^{1/2}$ has the form [9] $\langle \psi, (-\Delta + 1)^{1/2} \psi \rangle \geq K \int d^3x F(|\psi(x)|^2)$, where $F(a) = \int_0^a du (u^{2/3} + 1)^{1/2}$. One sees that for small a , $F(a)$ behaves like $ca^{5/3} + a$, which is similar to the $\rho_\psi^{5/3}$ bound one obtains for $\langle \psi, (-\Delta)\psi \rangle$ [2], and which is caused by the fact that for small p , $(p^2 + 1)^{1/2}$ behaves like $1 + p^2/2$. For large p , however, $(p^2 + 1)^{1/2}$ behaves linearly in p , which is reflected in the $a^{4/3}$ behavior of $F(a)$ for large a . Hence the lower bound on the kinetic energy is of the form $\int d^3x \rho_\psi(x)^{4/3}$ in the region where ρ_ψ is large; the corresponding Thomas–Fermi functional (including the other contributions to the energy) is not bounded below, and therefore does not lead to a useful lower bound on the ground state energy of $H_{N,K}$.

Heuristically, one can argue that Thomas–Fermi theories are “large Z theories” (i.e. they give the correct asymptotic behavior for $Z \rightarrow \infty$; this is rigorously true for the Schrödinger case [10].) Hence it is only natural that the Thomas–Fermi theory corresponding to $H_{N,K}$ is unbounded below, since $H_{N,K}$ itself is unbounded below whenever one Z_j becomes supercritical.

Throughout this paper we consider only the three-dimensional case, i.e. the Hilbert space we use is $L^2(\mathbb{R}^3)$. For functions on $\mathbb{R}^3$, the symbol $\| \cdot \|$ will be used to denote the $L^2(\mathbb{R}^3)$ -norm only; for operators, this symbol will always mean the norm as a bounded operator from $L^2(\mathbb{R}^3)$ to itself.

2. The Nuclear Repulsion Restores Stability

We shall prove Theorem 1 in two steps: first we shall prove it in the case $Z_k = Z_c$ all k . Then we shall apply a concavity argument to conclude the desired result for all Z_k . In the proof of the first step we shall need the following lemma, which can be considered as a refinement of the result $\| |x|^{-1/2} p |^{-1/2} \| = (\pi/2)^{1/2}$. This norm was evaluated in [5], and by a different method also in [11]; the critical value $Z_c e^2 = 2/\pi$ actually has its origin in this number.

Notation. $B(a, R) = \{x \mid |x - a| \leq R\}$.

Lemma 2.1. *Let ψ be an L^2 -function with $\text{supp } \psi \subset B(0, R)$. We denote by K the bounded operator $K = 2\pi^{-1}|x|^{-1/2}|p|^{-1}|x|^{-1/2}$. Then*

$$\langle \psi, K\psi \rangle \leq \langle \psi, \psi \rangle - \pi^{-3} R^{-2} \left(\int d^3x |x|^{-1/2} |\psi|^2 \right). \quad (2.1)$$

Proof. Note first that since ψ has compact support, the integral $\int d^3x |x|^{-1/2} |\psi(x)|$ converges. By simple scaling, we can obviously assume that $R = 1$, which we shall do henceforth in this proof.

We denote the symmetric decreasing rearrangement of ψ by ψ^* . Since $|x|^{-1/2}$ is symmetric and decreasing, we have $\int \psi^* |x|^{-1/2} \geq \int |\psi| |x|^{-1/2}$. Analogously, using the fact that $|p|^{-1}$ is the same as convolution with $(2\pi^2)^{-1} |x|^{-2}$, one sees that $\langle \psi^*, K\psi^* \rangle \geq \langle \psi, K\psi \rangle$ (this follows from the generalized rearrangement inequalities proved in [12].) Hence we can restrict our attention to ψ which are symmetric decreasing, i.e. $\psi = \psi^*$.

For any such $\psi(r)$, defined for $0 \leq r \leq 1$, we define, for $r \geq 1$, $\tilde{\psi}(r) = r^{-3} \psi(r^{-1})$. Consider the function $f(r) = \psi(r)$, $r \leq 1$, $f(r) = \tilde{\psi}(r)$, $r \geq 1$. One easily checks that f is in $L^2(\mathbb{R}^3)$, with $\|f\|^2 = 2\|\psi\|^2$. Straightforward calculation also leads to $\langle f, Kf \rangle = 2\langle \psi, K\psi \rangle + 2\langle \psi, L\psi \rangle$, where L has the integral kernel $L(x, y) = \pi^{-3} |x|^{-1/2} (1 + x^2 y^2 - 2x \cdot y)^{-1} |y|^{-1/2}$. Since $\| |x|^{-1/2} |p|^{-1/2} \| = (\pi/2)^{1/2}$, we have $\|K\| = 1$, hence $\langle f, Kf \rangle \leq \langle f, f \rangle$, which implies

$$\langle \psi, K\psi \rangle \leq \|\psi\|^2 - \langle \psi, L\psi \rangle. \quad (2.2)$$

By computing the spherical average of the kernel $L(x, y)$, we obtain

$$\begin{aligned} \langle \psi, L\psi \rangle &= (8/\pi) \int_0^1 dr \int_0^1 ds r^{1/2} \psi(r) s^{1/2} \psi(s) \ln[(1+rs)/(1-rs)] \\ &\geq (\pi^{-3}/2) \left[\int d^3x |x|^{-1/2} \psi(x) \right]^2 \min_{u \in [0, 1]} f(u), \end{aligned} \quad (2.3)$$

where $f(u) = u^{-1} \ln[(1+u)/(1-u)]$.

One can easily check that f attains its minimum at $u = 0$, with $f(0) = 2$. Expression (2.1) now follows immediately from (2.2) and (2.3). ■

Remark. The constant π^{-3} is probably not optimal in (2.1). We suspect that the optimal constant is given by functions behaving like $r^{-3/2}$. This is also the typical behavior of functions optimising the expectation value of $|x|^{-1/2} |p|^{-1} |x|^{-1/2}$ (see [11], [13]).

If we define $\psi_n(r) = 0$ for $r < n^{-1}$, $\psi_n(r) = r^{-3/2}$ for $n^{-1} < r \leq 1$, it turns out that

$$\lim_{n \rightarrow \infty} [\langle \psi_n, \psi_n \rangle - \langle \psi_n, K\psi_n \rangle] / \left[\int d^3x |x|^{-1/2} \psi_n(x) \right]^2 = a\pi^{-3},$$

with $a = \sum_{k=0}^{\infty} (2k+1)^{-3} = (7/8)\zeta(3) \simeq 1.05$. We think $a\pi^{-3}$ is probably the best constant for (2.1); in any event the sharp constant lies in the interval $[\pi^{-3}, 1.05\pi^{-3}]$.

With the help of Lemma 2.1, we can now prove Theorem 1 for the case where all the Z_j are critical.

Proposition 2.2. *For all $\psi \in Q(|p|)$, $\|\psi\| = 1$:*

$$\langle \psi, |p|\psi \rangle - (2/\pi) \sum_{j=1}^K \langle \psi, |x - R_j|^{-1} \psi \rangle \geq - (12/\pi) \sum_{\substack{j,k=1 \\ j < k}}^K |R_j - R_k|^{-1}. \quad (2.4)$$

Proof. Note first that since $|x|^{-1}$ is relatively form-bounded with respect to $|p|$ (see [4–6]), the left hand side of (2.4) makes sense for all ψ in $Q(|p|)$.

Let us rewrite the desired result (2.4) as

$$\forall \psi \in Q(|p|): \langle \psi, |p| \psi \rangle \geq (2/\pi) \sum_{j=1}^K \left\langle \psi, \left[|x - R_j|^{-1} - A \sum_{k \neq j} |R_k - R_j|^{-1} \right] \psi \right\rangle. \quad (2.5)$$

For the time being, we shall not fix A . We shall determine below a value A_0 such that (2.5) holds for all $A \geq A_0$. It will turn out that $A_0 = 3$, which then implies (2.4).

For r real, we shall use the notation $r_+ = \max(0, r)$. For K arbitrary real numbers $r_1, \dots, r_K$, we have

$$\sum_{j=1}^K r_j \leq \left(\sum_j r_j \right)_+ \leq \sum_j (r_j)_+ \leq \left(\sum (r_j)_+^{1/2} \right)^2.$$

Hence (2.5) will certainly be satisfied if

$$\forall \psi \in Q(|p|): \langle \psi, |p| \psi \rangle \geq \left\langle \psi, \left(\sum_{j=1}^K W_j(x) \right)^2 \psi \right\rangle, \quad (2.6)$$

where $W_j(x) = [(2/\pi)(|x - R_j|^{-1} - A \sum_{k \neq j} |R_k - R_j|^{-1})_+]^{1/2}$.

For any A , W_j has support in a ball with center R_j and radius t_j , where

$$t_j^{-1} = A \sum_{k \neq j} |R_k - R_j|^{-1}. \quad (2.7)$$

Provided A is large enough, all the balls $B(R_j, t_j)$ are disjoint: by (2.7), we have $t_j \leq A^{-1} |R_k - R_j|$ for all $k \neq j$. Therefore

$$|R_k - R_j| - (t_j + t_k) \leq (1 - 2/A) |R_k - R_j|. \quad (2.8)$$

which shows that $B(R_j, t_j) \cap B(R_k, t_k) = \emptyset$, for all $j \neq k$, if $A > 2$. From now on we shall assume $A > 2$.

For any j , let f_j be the characteristic function of $B(R_j, t_j)$. Since $W_j = W_j f_j$, and $W_j(x) \leq (2/\pi)^{1/2} |x - R_j|^{-1/2}$, we have

$$\sum_j W_j(x) = \sum_j W_j(x) f_j(x) \leq (2/\pi)^{1/2} \sum_j |x - R_j|^{-1/2} f_j(x) \equiv W(x). \quad (2.9)$$

Since $|x - R_j|^{-1/2} f_j \leq |x - R_j|^{-1/2}$, and $|x|^{-1/2} |p|^{-1/2}$ is bounded, one easily sees that $W|p|^{-1/2}$ is bounded. We shall prove now that as an operator from L^2 to L^2 , $\|W|p|^{-1/2}\| \leq 1$, provided A is larger than some constant which we shall evaluate below.

Let ψ be any function in C_0^∞ , so that $W\psi \in L^2$. Then

$$\begin{aligned} \| |p|^{-1/2} W\psi \|^2 &= (W\psi, |p|^{-1} W\psi) \\ &= (2/\pi) \sum_{j=1}^K (|x - R_j|^{-1/2} f_j \psi, |p|^{-1} |x - R_j|^{-1/2} f_j \psi) \\ &\quad + (2/\pi) \sum_{\substack{j,k=1 \\ j \neq k}}^K (f_j |x - R_j|^{-1/2} \psi, |p|^{-1} f_k |x - R_k|^{-1/2} \psi). \end{aligned} \quad (2.10)$$

The “diagonal” terms in (2.10) can be bounded above by applying Lemma 2.1:

$$\begin{aligned} & (2/\pi) \sum_{j=1}^K (f_j \psi, |x - R_j|^{-1/2} |p|^{-1} |x - R_j|^{-1/2} f_j \psi) \\ & \leq \sum_{j=1}^K \|f_j \psi\|^2 - \pi^{-3} \sum_{j=1}^K t_j^{-2} m_j^2, \end{aligned} \quad (2.11)$$

where

$$m_j = \int d^3x |x - R_j|^{-1/2} f_j |\psi|. \quad (2.12)$$

To find an upper bound for the “non-diagonal” terms in (2.10) we use (2.8) and the fact that $|p|^{-1}$ is the same as convolution with $(2\pi^2)^{-1} |x|^{-2}$. We obtain

$$\begin{aligned} & (2/\pi) (f_j |x - R_j|^{-1/2} \psi, |p|^{-1} f_k |x - R_k|^{-1/2} \psi) \\ & \leq \pi^{-3} [1 - 2/A]^{-2} |R_k - R_j|^{-2} m_j m_k, \end{aligned} \quad (2.13)$$

where m_j is defined by (2.12).

Combining (2.11) and (2.13) we obtain

$$\begin{aligned} \| |p|^{-1/2} W \psi \|^2 & \leq \| \psi \|^2 - \sum_{j=1}^K (\pi^{-3} t_j^{-2} m_j^2) \\ & + \pi^{-3} [1 - 2/A]^{-2} \sum_{\substack{j,k=1 \\ j \neq k}}^K m_j m_k |R_j - R_k|^{-2}. \end{aligned}$$

Using $m_j m_k \leq (m_j^2 + m_k^2)/2$ we can rewrite this as

$$\| |p|^{-1/2} W \psi \|^2 \leq \| \psi \|^2 - \sum_{j=1}^K m_j^2 b_j,$$

where $b_j = \pi^{-3} t_j^{-2} - \pi^{-3} [1 - 2/A]^{-2} \sum_{k \neq j} |R_k - R_j|^{-2}$. By the definition (2.7) of t_j , we have

$$b_j \geq \pi^{-3} [A^2 - (1 - 2/A)^{-2}] \sum_{k \neq j} |R_k - R_j|^{-2}.$$

This shows that all the b_j will be positive, and hence $\| |p|^{-1/2} W \psi \| \leq \| \psi \|$, if $A^2 \geq (1 - 2/A)^{-2}$, or

$$A \geq A_0 = 3. \quad (2.14)$$

Thus, provided (2.14) is satisfied, we have $\| |p|^{-1/2} W \psi \| \leq \| \psi \|$ for all ψ in C_0^∞ , hence $\| W |p|^{-1/2} \| = \| |p|^{-1/2} W \| = 1$. This implies $|p| - W^2 = |p|^{1/2} [1 - (W |p|^{-1/2})^* (W |p|^{-1/2})] |p|^{1/2} \geq 0$.

Since $\left(\sum_j W_j \right)^2 \leq W^2$ by (2.9), we have therefore proved (2.6) and *a fortiori* (2.5) for all $A \geq A_0 = 3$. ■

Remarks

1. In terms of Z_c and e^2 , we can rewrite (2.4) as $\forall \psi \in Q(|p|)$, $\|\psi\| = 1$:

$$\left\langle \psi, \left[|p| - Z_c \alpha \sum_{j=1}^K |x - R_j|^{-1} \right] \psi \right\rangle \geq -b Z_c^2 \alpha \sum_{\substack{j,k=1 \\ j < k}}^K |R_j - R_k|^{-1}, \quad (2.15)$$

where $b = \pi \alpha A_0 \leq 6.88 \times 10^{-2}$.

2. Note that (2.15) implies that $\sum_{j=1}^K Z_j \alpha |x - R_j|^{-1}$ is relatively form-bounded with respect to $|p|$, with relative bound $(\max Z_j)/Z_c$. This means that $h_{1,K}$ and $H_{1,K}$ can be defined as form sums, with form domain $Q(|p|)$ if $Z_j < Z_c$ for all j ; if $Z_j = Z_c$ for some j , we define $H_{1,K}, h_{1,K}$ to be the Friedrichs extensions of the quadratic forms on $Q(|p|)$.

With Proposition 2.2 ends our first step in the proof of Theorem 1 which is now established if $Z_j = Z_c$ for all j . From this result we shall derive Theorem 1 for all possible values of Z_j by a concavity argument. First we prove another technical lemma:

Lemma 2.3. *Let f be a real function on the n -dimensional unit cube $[0, 1]^n$, convex in each variable separately. Let h be the smallest concave function on $[0, 1]^n$ (jointly concave in all the variables) agreeing with f on all the vertices of the cube, i.e.*

$$\forall P_k \text{ vertex of } [0, 1]^n \quad (k = 1, \dots, M, \text{ with } M = 2^n): \quad h(P_k) = f(P_k).$$

Then $\forall (x_1, \dots, x_n) \in [0, 1]^n$: $h(x_1, \dots, x_n) \geq f(x_1, \dots, x_n)$.

Remarks

1. We shall apply this lemma to the function $f(x_1, \dots, x_n) = \sum_{i < j} \alpha_{ij} x_i x_j$. If we keep all the variables but one fixed, f is affine in the remaining variable. The requirement that f be convex in each variable separately is therefore obviously satisfied.

2. By changing signs, one immediately sees that the lemma is still true if we exchange “concave” and “convex,” if we make h the “largest convex function...” instead of the “smallest concave function...” and if we reverse the inequality sign in the conclusion.

3. Since h is the smallest concave function agreeing with f on the vertices, the conclusion $h \geq f$ automatically implies that h is the concave envelope of f . The lemma tells us therefore that to construct the concave envelope on a cube of a function which is convex in each variable separately, one only has to consider the values of f at the vertices of the cube.

4. At any point $(x_1, \dots, x_n)$ in the cube, h can be explicitly constructed as follows: for any decomposition of $(x_1, \dots, x_n)$ into a convex combination of the vertices: $x_m = \sum_{j=1}^M \lambda_j (P_j)_m$, $m = 1, \dots, n$, with $\lambda_j \geq 0$ and $\sum_{j=1}^M \lambda_j = 1$, we define $F(\underline{\lambda}) = \sum_{j=1}^M \lambda_j f(P_j)$; h is given by the maximum of $F(\underline{\lambda})$, taken over all possible convex decompositions of $(x_1, \dots, x_n)$.

Proof of Lemma 2.3. Fix $y_2, \dots, y_n$ so that each is either 0 or 1. Considered as a function of x_1 only, $(h-f)(x_1, y_2, \dots, y_n) = h(x_1, y_2, \dots, y_n) - f(x_1, y_2, \dots, y_n)$ is concave (h is jointly concave, and $-f$ is concave in each variable separately.)

Since h and f agree on the vertices, we have $(h-f)(0, y_2, \dots, y_n) \geq 0$, $(h-f)(1, y_2, \dots, y_n) \geq 0$. By the concavity of $h-f$ in x_1 , this implies $(h-f)(x_1, y_2, \dots, y_n) \geq 0$ for all $x_1 \in [0, 1]$. Now fix x_1 , and note that h and f satisfy the stated hypotheses in the variables $y_2, \dots, y_n$. The complete result follows by iterating the argument n times. ■

With the help of this lemma and Proposition 2.2, we can now prove Theorem 1:

Proof of Theorem 1. For $(z_1, \dots, z_K) \in [0, 1]^K$, we define

$$F(z_1, \dots, z_K) = \inf_{\psi \in Q(|p|), \|\psi\|=1} [\langle \psi, |p| \psi \rangle - (2/\pi) \sum_{k=1}^K z_k \langle \psi, |x - R_k|^{-1} \psi \rangle].$$

As the infimum of linear functions, F is concave. Define

$$f(z_1, \dots, z_K) = - (12/\pi) \sum_{\substack{k,l=1 \\ k < l}}^K |R_k - R_l|^{-1} z_k z_l. \quad (2.16)$$

Proposition 2.2 tells us that

$$F(P_k) \geq f(P_k) \quad k = 1, \dots, 2^K,$$

where the P_k , $k = 1, \dots, 2^K$ are the vertices of the cube $[0, 1]^K$. (When all the components of the vertices are 1, i.e. $P_k = (1, \dots, 1)$, (2.16) follows from Proposition 2.2. When some of the components of P_k are zero, (2.16) still follows from Proposition 2.2, but now with a smaller value for K .) Since F is concave, (2.16) implies that F is larger than h , the smallest concave function agreeing with f at the cornerpoints P_k of the cube. But h is larger than f by Lemma 2.3; by combining these two inequalities we obtain $F \geq f$, or for all $z_k \in [0, 1]$, $k = 1, \dots, K$:

$$\begin{aligned} & \inf_{\psi \in Q(|p|), \|\psi\|=1} [\langle \psi, |p| \psi \rangle - (2/\pi) \sum_{k=1}^K z_k \langle \psi, |x - R_k|^{-1} \psi \rangle] \\ & \geq - (12/\pi) \sum_{\substack{k,l=1 \\ k < l}}^K z_k z_l |R_k - R_l|^{-1}. \quad \blacksquare \end{aligned} \quad (2.17)$$

Remark. In terms of the fine-structure constant α and of the nuclear charges Z_k , with $Z_k \leq Z_c = 2/(\alpha\pi)$ for all k , (2.17) can be rewritten as

$$|p| - \sum_{k=1}^K Z_k \alpha |x - R_k|^{-1} \geq -b \sum_{\substack{k,l=1 \\ k < l}}^K Z_k Z_l \alpha |R_k - R_l|^{-1},$$

where $b \leq 6.88 \times 10^{-2}$.

3. Monotonicity of the Ground State Energy in the Nuclear Coordinates

We prove Theorem 2 by essentially the same method as used by one of us [8] in the proof of the analogous statement for the Schrödinger Hamiltonian

$-\Delta - \sum_{k=1}^K Z_k |x - R_k|^{-1}$. The strategy followed there was the following (modulo some epsilons):

$$1) E_0(\underline{R}) = \inf \text{spec } H(\underline{R}) = - \lim_{t \rightarrow \infty} t^{-1} \ln G(x, x, t, \underline{R}) \quad (3.1)$$

for all x , where $G(x, y, t, \underline{R})$ is the kernel $\exp[-tH(\underline{R})](x, y)$.

2) By the Trotter product formula, $G(x, y, t, \underline{R})$ can be approximated by multiple integrals involving $\exp[-(t/n)(-\Delta)](x, y)$ and $\exp[(t/n)Z_k|x - R_k|^{-1}]$. (Here one of the epsilons mentioned above comes into play: in order to ensure convergence of the integrals, one writes $G(x, y, t, \underline{R})$ as $\lim_{\varepsilon \rightarrow 0} G_\varepsilon(x, y, t, \underline{R})$,

where G_ε is the kernel of the semigroup corresponding to $-\Delta - \sum_{k=1}^K Z_k(|x - R_k|^2 + \varepsilon^2)^{-1/2}$. The whole argument is then carried out for G_ε rather than for G and the limit $\varepsilon \rightarrow 0$ is taken at the end.)

3) Since $\exp[as^{-1/2}]$ is completely monotone, it can be written as the Laplace transform of a positive measure; this can be used to write $\exp[(t/n)Z|x - R|^2]$ as an integral, over a positive measure, of $\exp[-s|x - R|^2]$.

4) Since all the kernels in the Trotter product approximation for G have been reduced to positive superpositions of Gaussian kernels (the kernel of $\exp[(t/n)\Delta]$ was already Gaussian in the Schrödinger case [8]), a lemma involving only Gaussian functions (Lemma 4 in [8]) can be applied, leading to the conclusion:

5) If $\underline{R}, \underline{R}'$ are given with $|R_i - R_j| \leq |R'_i - R'_j| \forall i, j$, and if x, y, x', y' are such that $|x - R_i| \leq |x' - R'_i|, |y - R_i| \leq |y' - R'_i|$ for all i , and $|x - y| \leq |x' - y'|$, then

$$G(x, y, t, \underline{R}) \geq G(x', y', t, \underline{R}').$$

Combined with 1) this proves the theorem. One uses the fact that given $\underline{R}, \underline{R}'$ and x , one can always find x' satisfying $|x' - R'_i| \geq |x - R_i|$ by choosing x' far enough away from the R'_i .

The same argument can be used to prove Theorem 2 in our case. Some modifications are needed, however. We shall only describe these in detail, and not repeat the whole proof as given in [8].

The first modification concerns the kernels $\exp[(t/n)(-\Delta + 1)^{1/2}]$ and $\exp[-(t/n)(-\Delta)^{1/2}]$. These are not Gaussian kernels, but they can again be written as positive superpositions of Gaussian kernels. Indeed, $\exp[-bs^{1/2}]$ is again a completely monotone function, and can therefore again, by Bernstein's theorem, be written as a Laplace transform of a positive measure:

$$\exp[-bs^{1/2}] = \int e^{-sv} d\mu(v, b);$$

an explicit formula for μ can be found in any table of Laplace transforms: $d\mu(v, b) = (b\pi^{-1/2}/2)v^{-3/2} \exp[-b^2/4v]dv$. It follows now that

$$\begin{aligned} \exp[-(t/n)(-\Delta + 1)^{1/2}](x, y) &= \int (2v)^{-3/2} \exp[-|x - y|^2/4v] e^{-v} d\mu(v, t/n), \\ \exp[-(t/n)(-\Delta)^{1/2}](x, y) &= \int (2v)^{-3/2} \exp[-|x - y|^2/4v] d\mu(v, t/n), \end{aligned} \quad (3.2)$$

which are indeed positive superpositions of Gaussians.

The second modification is more a set of comments on step 1) than a real modification.

First of all, note that it is enough to prove Theorem 2 for the case $Z_k < Z_c$ for all k . As the infima of decreasing linear functions of the Z_k , $E_{1,k}^0(Z)$ and $e_{1,k}^0(Z)$ are decreasing, upper semicontinuous functions of the Z_k , which ensures that

$$E_{1,K}^0(Z) = \lim_{\varepsilon \rightarrow 0} E_{1,K}^0(Z - \varepsilon), e_{1,K}^0(Z) = \lim_{\varepsilon \rightarrow 0} e_{1,K}^0(Z - \varepsilon),$$

where we use the notation $Z - \varepsilon$ for the set $\{Z_j - \varepsilon\}$. This implies that if Theorem 2 holds for the case where $Z_k < Z_c$ for all k , it also holds in the case where some of the Z_k are equal to Z_c .

We can therefore restrict ourselves to the case $Z_k < Z_c$, all k . In this case $E_{1,K}^0$ is an isolated eigenvalue (one can easily show $E_{1,K}^0 < 1$ by a variational argument, and since $Z_k < Z_c$ for all k , we know that $\sigma_{\text{ess}}(H_{1,K}^0) = [1, \infty)$). For $e_{1,K}^0$, matters are slightly different. If $\sum_{k=1}^K Z_k \leq Z_c$, we know that $e_{1,K}^0(Z, R) = 0$, independently of R , in which case the monotonicity in R is trivial. We shall therefore not consider this case in the discussion below. Whenever $\sum_{k=1}^K Z_k > Z_c$, however, one has again $e_{1,K}^0 < 0$ by a variational argument, and since $\sigma_{\text{ess}}(h_{1,K}^0) = [0, \infty)$, this shows that then $e_{1,K}^0$ too is an isolated eigenvalue. We shall therefore assume in the following that $E_{1,K}^0, e_{1,K}^0$ are isolated eigenvalues. We know then that (3.1) will be true, for all x , if there exist ground states for $H_{1,K}^0, h_{1,K}^0$ which are strictly positive. The argument in the next paragraph shows that this is indeed the case: we show that the ground states for $H_{1,K}^0, h_{1,K}^0$ (where $\sum_{k=1}^K Z_k > Z_c$ in the last case, and $Z_k < Z_c$ for all k for both Hamiltonians) are

nondegenerate and strictly positive, which is an interesting result in its own right. Since $H_{1,K}^0 = H_0 - V(R, x)$, $h_{1,K}^0 = h_0 - V(R, x)$, where $H_0 = (-\Delta + 1)^{1/2}$, $h_0 = (-\Delta)^{1/2}$, and where V is positive, the kernels $\exp[-tH_{1,K}^0](x, y)$, $\exp[-th_{1,K}^0](x, y)$ will be pointwise larger than $\exp[-tH_0](x, y)$ and $\exp[-th_0](x, y)$, respectively. By (3.2) one sees immediately that $\exp[-tH_0](x, y)$, $\exp[-th_0](x, y)$ are strictly positive for all x, y , which implies therefore $\exp[-tH_{1,K}^0](x, y) > 0$, $\exp[-th_{1,K}^0](x, y) > 0$ for all x, y . By the extension of the Perron–Frobenius theorem to operators (see e.g. [14], Theorem XIII.43), this implies that the ground states of $H_{1,K}^0, h_{1,K}^0$ (which correspond to the largest eigenvalues of $\exp[-tH_{1,K}^0]$, $\exp[-th_{1,K}^0]$) are nondegenerate, and the corresponding eigenvectors strictly positive.

Taking into account all the remarks made above, the proof given in [8] can now be completely transcribed to our case; this completes our proof of Theorem 2.

References

1. Dyson, F., Lenard, A.: Stability of matter. I. J. Math. Phys. **8**, 423–434 (1967)
Lenard, A., Dyson, F.: Stability of matter. II. J. Math. Phys. **9**, 698–711 (1968)
2. Lieb, E., Thirring, W.: A bound for the kinetic energy of fermions which proves the stability of matter.

- Phys. Rev. Lett. **35**, 687–689 (1975); Errata: Phys. Rev. Lett. **35**, 1116 (1975). See also Lieb, E.: The stability of matter. Rev. Mod. Phys. **48**, 553–569 (1976)
3. Dyson, F.: Ground-state energy of a finite system of charged particles. J. Math. Phys. **8**, 1538–1545 (1967)
 - Lieb, E.: The $N^{5/3}$ law for bosons. Phys. Lett. **A70**, 71–73 (1979)
 4. Weder, R.: Spectral analysis of pseudodifferential operators. J. Funct. Anal. **20**, 319–337 (1975)
 5. Herbst, I.: Spectral theory of the operator $(p^2 + m^2)^{1/2} - Ze^2/r$. Commun. Math. Phys. **53**, 285–294 (1977); Errata: Commun. Math. Phys. **55**, 316 (1977)
 6. Kato, T.: Perturbation theory for linear operators. Berlin, New York: Springer 1966 (2nd edn. 1976)
 7. Lieb, E., Simon, B.: Monotonicity of the electronic contribution to the Born–Oppenheimer energy. J. Phys. **B11**, L537–542 (1978)
 8. Lieb, E.: Monotonicity of the molecular electronic energy in the nuclear coordinates. J. Phys. **B15**, L63–L66 (1982)
 9. Daubechies, I.: An uncertainty principle for fermions with generalized kinetic energy. Commun. Math. Phys. (1983)
 10. Lieb, E., Simon, B.: The Thomas–Fermi theory of atoms, molecules and solids. Adv. Math. **23**, 22–116 (1977)
 11. Kovalenko, V., Perelmuter, M., Semenov, Ya.: Schrödinger operators with $L_w^{1/2}(\mathbb{R}^d)$ potentials. J. Math. Phys. **22**, 1033–1044 (1981)
 12. Brascamp, H., Lieb, E., Luttinger, M.: A general rearrangement inequality for multiple integrals. J. Funct. Anal. **17**, 227–237 (1974)
 13. Lieb, E.: Sharp constants in the Hardy–Littlewood–Sobolev and related inequalities (submitted)
 14. Reed, M., Simon, B.: Methods of modern mathematical physics Vol. IV: Analysis of operators. New York: Academic Press 1978

Communicated by J. Fröhlich

Received March 24, 1983

The Stability and Instability of Relativistic Matter

Elliott H. Lieb^{1,*} and Horng-Tzer Yau^{2,**}

¹ Departments of Mathematics and Physics, Princeton University, P.O. Box 708,
Princeton, NJ 08544, USA

² School of Mathematics, The Institute for Advanced Study, Princeton, NJ 08540, USA

Abstract. We consider the quantum mechanical many-body problem of electrons and fixed nuclei interacting via Coulomb forces, but with a relativistic form for the kinetic energy, namely $p^2/2m$ is replaced by $(p^2c^2 + m^2c^4)^{1/2} - mc^2$. The electrons are allowed to have q spin states ($q=2$ in nature). For one electron and one nucleus instability occurs if $z\alpha > 2/\pi$, where z is the nuclear charge and α is the fine structure constant. We prove that stability occurs in the many-body case if $z\alpha \leq 2/\pi$ and $\alpha < 1/(47q)$. For small z , a better bound on α is also given. In the other direction we show that there is a critical α_c (no greater than $128/15\pi$) such that if $\alpha > \alpha_c$ then instability always occurs for *all* positive z (not necessarily integral) when the number of nuclei is large enough. Several other results of a technical nature are also given such as localization estimates and bounds for the relativistic kinetic energy.

I. Introduction

One of the early important successes of quantum mechanics was the interpretation of the stability of the hydrogen atom. The ground state energy of the hydrogen Hamiltonian is finite and thus the hydrogen atom is stable quantum mechanically, even though it is unstable classically. The Coulomb singularity $-ze^2/r$ is controlled by a new feature of Schrödinger mechanics, the uncertainty principle. While the stability of the hydrogen atom is clear and simple, a more subtle question arises when many particles are taken into account. It is convenient to distinguish two notions of stability.

Stability of the first kind: The ground state energy is finite.

Stability of the second kind: The ground state energy is bounded below by a constant times the number of particles.

* Work partially supported by U.S. National Science Foundation grant PHY-85-15288-A02

** The author thanks the Institute for Advanced Study for its hospitality and the U.S. National Science Foundation for support under grant DMS-8601978

The second kind of stability, now commonly known as the stability of matter, was proved in 1967 by Dyson and Lenard [10] – four decades after the invention of Schrödinger mechanics. The Dyson-Lenard analysis clearly showed that the stability of matter depends crucially on the Pauli exclusion principle. The ground state energy (call it E_f) of N fermions interacting with K infinitely massive nuclei via the Coulomb potential is bounded below by a constant times the total particle number, i.e. $E_f \geq -C_1(N+K)$. On the other hand, if all the particles considered are bosons, Dyson and Lenard [10] showed that the ground state energy (call it E_b) satisfies $E_b \geq -C_2(N+K)^{5/3}$. Lieb [20] showed that this $5/3$ bound is indeed the correct law for infinitely massive nuclei. If the nuclei have finite mass, and are also bosons, Dyson [9] showed by a variational calculation, that the ground state energy of bosons is bounded above by $E_b \leq -C_3(N+K)^{7/5}$. This clearly shows that bosons are stable in the first sense, but never in the second. Dyson [9] also conjectured a lower bound $E_b \geq -C_4(N+K)^{7/5}$ and this was finally proved 20 years later by Conlon, Lieb, and Yau [4]. They also proved a related bound for bosonic jellium.

The Dyson-Lenard proof for fermions involved a sequence of inequalities such that the final bound for C_1 is 10^{14} Rydberg. New proofs were given by Federbush [12] and Lieb-Thirring [25] in the seventies. The Lieb-Thirring proof gave a much better bound on C_1 (23 Rydbergs) and related the stability problem to the semiclassical picture of Thomas-Fermi theory. These matters are reviewed in [19].

The aforementioned considerations are all based on the nonrelativistic Schrödinger equation. The kinetic energy operator is the standard $p^2/2m = -\Delta/2m$ (when $\hbar=1$). One might wonder whether stability still prevails in the relativistic case since the kinetic energy then decreases from $p^2/2m$ to $(p^2+m^2)^{1/2} - m$ ($\hbar=c=1$). Historically, Chandrasekhar [2] was one of the first to ask this question, but in the context of gravitational interaction instead of Coulomb interaction. The famous Chandrasekhar model for neutron stars or white dwarfs consists of a semiclassical relativistic kinetic energy and classical gravitational potential energy. This simple model remarkably predicted collapse (i.e. instability of the first kind) and gave a critical mass which is correct, at least approximately. Despite the success of the simple semi-relativistic Chandrasekhar theory, the kinetic energy operator,

$$T = (p^2 + m^2)^{1/2} - m,$$

which it employs is nonlocal and therefore violates a basic physical principle. Nevertheless it is worthwhile studying this operator for several reasons. When $m=0$, $T=|p|$ and it has the correct inverse length scaling (like the Dirac operator). Unlike the Dirac operator it allows one to formulate a variational principle for the ground state energy and thereby to give a rigorous definition of stability without the necessity of filling the Dirac sea or invoking quantum electrodynamics. In any event, there does not exist a truly relativistic many-body quantum theory at the present time and it is our belief that the study of Schrödinger operators based on T will capture some of the essential features of “the correct theory” when it is eventually formulated.

Let us start with the Hydrogen atom by considering the one particle Hamiltonian $\tilde{H}$ defined by

$$\tilde{H}_1 = (p^2 + m^2)^{1/2} - m - \alpha z/|x|, \quad (1.1)$$

where $\alpha = e^2$ is the fine structure constant ($\hbar = c = 1$). This operator was studied independently by Weder [29] and Herbst [16]. See also Daubechies' paper [7]. Since the difference between the operator $(p^2 + m^2)^{1/2} - m$ and $|p|$ is bounded (more precisely $|p| \geq (p^2 + m^2)^{1/2} - m \geq |p| - m$), the stability of (1.1) is the same as the stability of

$$H_1 = |p| - \frac{2}{\pi} \beta/|x|, \quad (1.2)$$

where

$$\beta = \pi \alpha z/2. \quad (1.3)$$

Note that (1.2) is homogeneous under length scaling and therefore $E_1 \equiv \inf \text{spec } H_1$ is either 0 or $-\infty$ by the scaling $\psi(x) \rightarrow \lambda^{3/2} \psi(\lambda x)$.

A first important fact about (1.2) is the existence of a critical $\beta_c = 1$, similar to that of the Klein-Gordon or Dirac theories. Kato [17] stated that $\beta_c \geq 1$ and Herbst [16] showed that $\beta_c = 1$. The ground state energy for the Hamiltonian (1.2) is $E_1 = -\infty$ if $\beta > 1$ and $E_1 = 0$ if $\beta \leq 1$. (In the Dirac theory $\beta_c = \pi/2$.)

Returning to the many-body case, suppose we have N electrons with coordinates $x_1, \dots, x_N$ in $\mathbb{R}^3$ and K nuclei with coordinates $R_1, \dots, R_K$ in $\mathbb{R}^3$ and with positive charges $z_1, \dots, z_K$. We shall consider the following relativistic Schrödinger Hamiltonian, H_{NK} , for fermions with q spin states ($q=2$ for real electrons). It is the analogue of (1.2):

$$H_{NK} \equiv \sum_{i=1}^N |p_i| + \alpha V_c(x_1, \dots, x_N; R_1, \dots, R_K), \quad (1.4)$$

$$\begin{aligned} V_c(x_1, \dots, x_N; R_1, \dots, R_K) \equiv & \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} - \sum_{i=1}^N \sum_{j=1}^K z_j |x_i - R_j|^{-1} \\ & + \sum_{1 \leq i < j \leq K} z_i z_j |R_i - R_j|^{-1}. \end{aligned} \quad (1.5)$$

Note that *charge neutrality is not assumed in (1.4), or anywhere else in this paper.*

Mathematically, the Hamiltonian H_{NK} is a quadratic form on the q -state physical subspace $\mathcal{H}^q$ of $L^2(\mathbb{R}^{3N})$. More precisely, $\psi \in \mathcal{H}^q$ if and only if there exists a partition $P = \{\pi_1, \dots, \pi_q\}$ of $\{1, \dots, N\}$ such that $\psi(x_1, \dots, x_N)$ is an antisymmetric function of the variables in each π_j , for all $1 \leq j \leq q$. When $q=N$, there is no restriction and the ground state energy for H_{NK} is just the ground state energy for bosons.

Physically, the nuclear kinetic energies should be included in (1.4) since the Born-Oppenheimer approximation (i.e. the neglect of the nuclear kinetic energies) is inadequate in the extreme relativistic regime. For simplicity, we shall confine ourselves to the Born-Oppenheimer approximation.

In reality, our goal is to discuss stability of the second kind for $H_{NK}(m)$, which is given by (1.4) but with $|p|$ replaced by $(p^2 + m^2)^{1/2} - m$ there. For this purpose,

however, it suffices to study only stability of the first kind for H_{NK} in (1.4). The reason is the following. Let $E_{NK}(R_1, \dots, R_K)$ denote the ground state energy ($= \inf \text{spec}$) of H_{NK} and let E_{NK} be the infimum of $E_{NK}(R_1, \dots, R_K)$ over all choices of the R 's. By simple scaling ($\psi(x_1, \dots, x_N) \rightarrow \lambda^{3N/2} \psi(\lambda x_1, \dots, \lambda x_N)$ and $R_j \rightarrow R_j/\lambda$), we see that E_{NK} is either zero or $-\infty$. On the other hand, if $E_{NK}(m)$ is defined analogously, then, since $|p| - m < (p^2 + m^2)^{1/2} - m < |p|$, we have that $E_{NK} \geq E_{NK}(m) \geq E_{NK} - mN$. Thus *stability of the first kind for H_{NK} (in the sense that E_{NK} is bounded below independent of the R_j) is equivalent to stability of the second kind for $H_{NK}(m)$* . Our goal then – and that is the purpose of this paper – is to find necessary conditions and sufficient conditions on z and α so that $E_{NK}(R_1, \dots, R_K) \geq 0$ for all N and all K and all $R_1, \dots, R_K$.

If everything is held fixed except for q , then $E_{NK}(R_1, \dots, R_K)$ is a monotone decreasing function of q . The reason is that specifying q is the same thing as requiring that the admissible wave functions $\psi(x_1, \dots, x_N)$ are antisymmetric in each of q sets of variables. The number of variables in each set is unimportant, zero being an allowed number. Thus, a valid function for q is trivially a valid function for $q+1$.

A further remark about (1.4) can be made. Using a convexity argument, Daubechies and Lieb [8] proved that the stability of H_{NK} for $z_1 = z_2 = \dots = z_K = z$ implies the stability of H_{NK} when all the nuclear charges are no greater than z , i.e. $0 \leq z_j \leq z$ for all j . With this remark, we shall assume from now on $z_1 = \dots = z_K = z$.

Let $E_{NK}(\alpha, z)$ denote the dependence of E_{NK} on α and z . We shall use the following terminology: $H(\alpha, z)$ is *stable* means that $E_{NK}(\alpha, z) = 0$ for all N and K . Otherwise we say that $H(\alpha, z)$ is *unstable*.

The coupling constant of the electrons to the nuclei is $z\alpha = 2\beta/\pi$ and, from the hydrogen atom result, it is clearly necessary to have $\beta \leq 1$ for stability. It is frequently convenient, therefore, to adopt α and β as the independent variables instead of α and z . When doing so we shall refer to the stability or instability of $H(\alpha, \beta)$ – hopefully without confusion. Indeed α and β are the natural variables from the following point of view. The electron-nuclear coupling is $2\beta/\pi$ while the nuclear-nuclear repulsion constant is $z^2\alpha = (2/\pi)^2\beta^2/\alpha$. Suppose that $K > 1$ and $\beta < 1$, but $K\beta > 1$. Then, if the nuclear-nuclear repulsion is ignored, the K nuclei can come to one common point and the system will collapse – even with only one electron. What discourages this from happening is the repulsion which is proportional to β^2/α . With β fixed, we see that α is required to be small in order that this repulsion prevents collapse. It is a striking fact, and it is the main theme of this paper, that *for every fixed $\beta \leq 1$ and q there is a critical α (call it $\alpha_c(\beta)$) so that $H(\alpha, \beta)$ is stable when $\alpha < \alpha_c(\beta)$. There is another critical α (call it $\bar{\alpha}_c(\beta)$) so that $H(\alpha, \beta)$ is unstable when $\alpha > \bar{\alpha}_c(\beta)$* . These facts are the reason behind the contention above that α and β are natural. We do not know whether or not $\alpha_c(\beta) = \bar{\alpha}_c(\beta)$. Note that by the above monotonicity in z remark, stability for some (α, β_1) implies stability for all (α, β) with $\beta < \beta_1$.

There is an additional piece of information. Suppose that stability occurs for a pair α_1, z . Then stability occurs for a pair α, z if $\alpha \leq \alpha_1$. The reason for the monotonicity in α is that $\inf \text{spec}(\sum |p_i| + \alpha V_C) \geq (1 - \alpha/\alpha_1) \inf \text{spec}(\sum |p_i|) + (\alpha/\alpha_1) \inf \text{spec}(\sum |p_i| + \alpha_1 V_C) \geq 0$.

Before stating our main results in detail, let us review some recent progress with this and related problems that also have the feature of critical coupling constants.

(1) The Chandrasekhar critical mass was established up to a factor of 4 in the framework of the relativistic Schrödinger equation by Lieb-Thirring [26]. Later, Lieb-Yau [27] proved that not only is the Chandrasekhar critical mass exactly correct, but the Chandrasekhar semiclassical equation can be derived rigorously from the relativistic Schrödinger equation in the limit that the gravitational constant $G \rightarrow 0$. In particular, in the physically interesting case, the discrepancy between the Chandrasekhar semiclassical critical mass and the quantum mechanical critical mass was shown in [27] to be less than 0.01%.

(2) For the non-relativistic Schrödinger equation, but with magnetic fields present that couple to both the electronic orbital motion and electronic spin, the existence of a critical nuclear charge for the stability of the hydrogen atom was proved by Fröhlich, Lieb, Loss, and Yau [15, 28]. The results were extended to the one-electron molecule and many-electron atom by Lieb and Loss [23]. The stability criteria are very similar to that of the relativistic stability considered in this paper. For stability, one should keep *both* $\alpha^2 z$ and α small. The general case for this model (many electrons and nuclei) remains an interesting open problem.

(3) The relativistic stability of matter itself. For $N=1$ and K arbitrary, Daubechies and Lieb [8] were the first to note the existence of a critical α and β fixed. They proved that H_{1K} is stable in the critical case $\beta = \pi\alpha z/2 = 1$ if $\alpha \leq 1/3\pi$. The first person to solve a general case for all N and K was Conlon [3], who proved that the Hamiltonian $H(\alpha, z)$ is stable when $z=1$ provided $\alpha \leq 10^{-200}$ and $q=1$. Using a different method, Fefferman and de la Llave [14] improved Conlon's result for $z=1$ to $\alpha \leq 1/2.06\pi$, and again $q=1$. The Fefferman-de la Llave proof used computer assisted proofs extensively. Without using a computer, their bound would be worse by a factor 2.5, thereby reducing α to $1/5\pi$. Recently, Fefferman [13] announced a result for the critical case $\beta=1$ provided some numerical computer calculations can be made rigorous. The stability criterion announced in [13] is that stability occurs in the critical case $\beta=1$ if $\alpha \leq 1/20$ and $q=1$. A complete proof, however, was not available when the present paper was written. Since $H(\alpha, \beta)$ collapses for $\beta > 1$ no matter how small the difference $\beta - 1$ may be, the application of computer assisted proofs to the $\beta=1$ case is delicate and difficult. Fefferman [13] states that "arbitrarily small roundoff errors are apparently fatal."

All the results mentioned above address the situation $q=1$. The methods employed are not, in our opinion, easily generalized to treat arbitrary q , as is done here. The ability to treat arbitrary q without increasing the complexity of the proof as q increases is, in our opinion, one of the main advantages of our method. Another is that we have no intrinsic need to invoke the computer. *The essence of our method is that for all q the many-body problem is reduced to a tractable one-body problem (see e.g. Theorems 6 and 11). This method also makes it possible to prove, for the first time, that stability occurs up to and including the critical value $\beta=1$.*

We should point out that the main tool in proving the nonrelativistic stability of matter, the Thomas-Fermi theory, fails to predict stability in the relativistic case. The semiclassical kinetic energy decreases in the high momentum region from $(\text{const}) \int \varrho^{5/3}$ in the nonrelativistic case to $(\text{const}) \int \varrho^{4/3}$ in the relativistic case. This semiclassical kinetic energy, $\int \varrho^{4/3}$, cannot control the Coulomb singularity $z\alpha/r$ for

any $\alpha > 0$. The fact that stability occurs only for some finite $\alpha > 0$ and $z > 0$ is not a trivial matter (see Conlon [3]). A good estimate for α , especially when β is set equal to its critical value 1, is very difficult to achieve and should resolve the following subtle points:

(i) The delicate balance of charge neutrality. If, for example, the attractive term in V_c is changed from $z\alpha \sum \sum |x_i - R_j|^{-1}$ to $z\alpha(1+\varepsilon) \sum \sum |x_i - R_j|^{-1}$ for some $\varepsilon > 0$, then stability will not occur for any positive α and z . Physically, an attractive gravitational interaction is present and it does alter the Hamiltonian in precisely this manner – collapse does indeed occur. But the gravitational constant is small, and this collapse happens only when N and K are extremely large – the order of a solar mass [26, 27]. Indeed, the problem of determining the critical mass when Coulomb and gravitational interactions are both taken into account is a difficult open problem.

(ii) An improved version of the basic inequality $|p| - \frac{2}{\pi} |x|^{-1} \geq 0$ is needed. This is apparently crucial since each electron in general feels attractions from more than one nucleus. One may argue that, by virtue of screening, each electron feels only one attraction from its nearest nucleus, but it is difficult to find a simple, precise mathematical statement about screening. Indeed, some corrections (e.g. van der Waals force) are obviously unavoidable and can only be controlled by the kinetic energy.

(iii) The nonlocality of the operator $|p|$. The technical problems caused by this non-locality are serious, especially since the Coulomb potential is long-ranged.

Our main results are the following four theorems about stability and instability.

Theorem 1 (Simple Stability Criterion). *For any $z > 0$ and q , the Hamiltonian $H(\alpha, z)$ is stable if*

$$\alpha \leq \sup_{z' \geq z} A_q(z'), \quad (1.6)$$

where

$$A_q(z) = (2/\pi) z^{-1} [1 + q^{1/3} z^{-1/3} C(z)^{-1/3}]^{-1}, \quad (1.7)$$

$$C(z) = 3.0844 \{ [1.6617 + 1.7258 z^{-1} + 0.9533 z^{-1/2}]^4 + (4/\pi)^3 [1 + (2z)^{-1/2}]^8 \}^{-1}. \quad (1.8)$$

Corollary. Fix $\beta \equiv z\alpha\pi/2 < 1$. Then stability occurs if

$$q\alpha \leq \begin{cases} 0.062980(1-\beta)^3\beta^{-2} & \text{if } \beta \geq 0.49910 \\ 0.031774 & \text{if } \beta \leq 0.49910. \end{cases} \quad (1.9)$$

Remark. There is a number z_1 , which is roughly 0.6, such that if $z \geq z_1$ then the supremum in (1.6) occurs for $z' = z$, while if $z \leq z_1$ the supremum occurs for $z' = z_1$.

Theorem 2 (Stability criterion for $\beta \leq 1$). Fix $\beta \leq 1$. Then the Hamiltonian $H(\alpha, \beta)$ is stable if

$$q\alpha \leq 1/47.$$

Theorem 3 (Instability for all z and q). There is a critical value α_1 such that if $\alpha > \alpha_1$ then $H(\alpha, z)$ is unstable for every $q \geq 1$ and every nuclear charge $z > 0$ (not necessarily

integral), no matter how small z may be. This means that if $\alpha > \alpha_1$, one can always choose N and K so that $E_{NK}(\alpha, z) = -\infty$. In order to achieve this collapse, it is only necessary to use one electron, i.e. $N=1$. One can take $\alpha_1 = 128/15\pi$.

Theorem 4 (Instability dependence on q). *Let $\beta = \pi\alpha z/2$ as in (1.3). There is a critical value α_2 such that if*

$$\alpha > \alpha_2 q^{-1} \beta^{-2}, \quad (1.10)$$

then $H(\alpha, \beta)$ is always unstable. To achieve this collapse, only $N=q$ electrons are needed. One can take $\alpha_2 = 115, 120$. Alternatively, $\alpha > 36q^{-1/3} z^{2/3}$.

Corollary. *If the electrons are bosons then $H(\alpha, z)$ is unstable for all $\alpha > 0$ and all fixed $z > 0$. The number of electrons necessary to achieve this collapse satisfies $N \leq 4\pi^{-2} \alpha_2 z^{-2} \alpha^{-3}$.*

Remarks. In view of Theorem 3, the number 115, 120 should not be taken seriously. Its large value merely demonstrates how difficult it is to find simple, rigorous bounds – even variational upper bounds – for the relativistic Coulomb problem.

These theorems, taken together, give a clear picture about the stability of relativistic matter. The relevant parameters for stability are αq (if β is fixed) and $\alpha q^{1/3}$ (if z is fixed). An upper bound for α which is independent of z and q is given in Theorem 3. β is never larger than 1. Theorem 1 clearly fails to predict stability for the critical case $\alpha z = 2/\pi$, but its proof is considerably simpler than that of Theorem 2. It also gives the correct q dependence (when z is fixed), and its bound on α for small z is better than that of Theorem 2.

To gain perspective on how good these bounds are, we specialize our results to the following two cases. First, in the critical case, our upper bound (Theorem 2) and lower bound (Theorem 3) differ by a factor of 128 for $q=1$. Second, for $z=1$ and $q=1$, Theorem 1 predicts stability for $\alpha \leq 1/3.23\pi$, which is not appreciably worse than the computer assisted proof bound $1/2.06\pi$ in [14]. Our bounds in Theorem 1 and Theorem 2 can certainly be improved, as will become clear in the proofs given below. We refrain from the temptation to optimize our results by complicating the technicalities. Our goal is to give a simple conceptual proof which has the correct q dependence and reasonable estimates.

Our proofs for Theorem 3 and 4 follow the same idea used in [23, 20]. Theorems 1 and 2 are much more difficult. Our basic strategy is first to reduce the Coulomb potential to a one-body potential, W . Then, by localizing the kinetic energy $|p|$, we can control the short distance Coulomb singularity of W , leaving a bounded potential W^* as remainder. The last task is to bound the sum of the negative eigenvalues of $|p| + W^*$, but this is standard and can be done by using semiclassical bounds ([6]).

The following Theorem 5 is a consequence of our localization for $|p|$ and combinatorial ideas in [26]. Theorem 5 was announced in [27, Appendix B], where it was proved for the special case $q=N$. Earlier, Fefferman and de la Llave [14] proved it for $q=1$. This theorem is not needed in the present work, but it is independently interesting. (Note that the definition of δ_i below is the reciprocal of that in [27].)

Theorem 5 (Domination of the nearest neighbor attraction by kinetic energy). *Let $\delta_i = \delta_i(x_1, \dots, x_N)$ be the nearest neighbor distance for particle i relative to $N - 1$ other particles, i.e.*

$$\delta_i \equiv \min \{ |x_i - x_j| | j \neq i \}. \quad (1.11)$$

Let $\psi \in L^2(\mathbb{R}^{3N})$ be an N particle fermionic function of space-spin with q spin states. Then

$$\sum_{i=1}^N (\psi, |p_i| \psi) \geq C_1 q^{-1/3} \sum_{i=1}^N (\psi, \delta_i^{-1} \psi), \quad (1.12)$$

$$\sum_{i=1}^N (\psi, p_i^2 \psi) \geq C_2 q^{-2/3} \sum_{i=1}^N (\psi, \delta_i^{-2} \psi), \quad (1.13)$$

where

$$C_1 = 0.129, \quad C_2 = 0.0209. \quad (1.14)$$

The organization of the rest of this paper is as follows:

In Sects. II and III, we prove Theorems 1 and 2 assuming an electrostatic inequality for the Coulomb potential and localization estimates for $|p|$. The theorems used in Sects. II and III are then proved in Sects. IV–VII. The presentation has been broken up this way in order to stress the conceptual underpinnings of Theorems 1 and 2.

Theorem 5 is proved in Sect. V. Some details of our numerical calculations are explained in Sect. VIII. In the final Sect. IX we prove Theorems 3 and 4.

II. Proof of Theorem 1 ($z\alpha < 2/\pi$)

The proofs of Theorems 1 and 2 are conceptually much simpler than the following detailed calculations and technicalities would suggest. There are three main steps for Theorem 1 and five steps for Theorem 2. Step A is the same for both theorems.

Step A. Reduction of the many-body Coulomb potential to a sum of one-body potentials plus a positive constant, namely $-\sum_{i=1}^N W(x_i) + C$. This reduces the problem to that of showing that q times the sum of the negative eigenvalues of the operator $|p| - W$ is not less than $-C$.

In the next step we decompose $\mathbb{R}^3$ into regions $B_0, B_1, \dots, B_K$ where the B_i are disjoint balls centered at the R_i and B_0 is everything else.

Step B. We write $|p| = \beta|p| + (1 - \beta)|p|$ with $\beta = z\alpha\pi/2 < 1$. In the balls $B_i, i = 1, \dots, K$ we use $\beta|p|$ to control the Coulomb singularity of W and prove the operator inequality

$$\beta|p| - \alpha W(x) \geq -U(x), \quad (2.1)$$

where $U = W$ in B_0 and U is a continuous function inside each ball. Thus $|p| - \alpha W \geq (1 - \beta)|p| - U$.

Step C. The sum of the negative eigenvalues of $(1 - \beta)|p| - U$ is bounded by using the semiclassical bound due to Daubechies [6].

Steps B, C, D, and E for Theorem 2 will be explained in Sect. III.

In this section we shall state the basic theorems for steps A and B. These will be proved later in Sects. IV and V. These theorems will be combined here in step C, thus completing the proof of Theorem 1.

Step A. Reduction of the Coulomb Potential to a One-Body Potential

This step has nothing to do with quantum mechanics or the nature of the kinetic energy operator. It has to do with screening in classical potential theory. The total Coulomb potential, V_c , is given in (1.5). There are K nuclei located at distinct points $R_1, \dots, R_K$ in $\mathbb{R}^3$ and having the same charge, z . There are N electrons.

Introduce the nearest neighbor, or Voronoi, cells $\{\Gamma_j\}_{j=1}^K$ defined by

$$\Gamma_j = \{x \mid |x - R_j| \leq |x - R_k| \text{ for all } k \neq j\}. \quad (2.2)$$

The boundary of Γ_j , $\partial\Gamma_j$, consists of a finite number of planes. Another important quantity is the distance

$$D_j = \text{dist}(R_j, \partial\Gamma_j) = \frac{1}{2} \min \{|R_k - R_j| \mid j \neq k\}. \quad (2.3)$$

The following theorem will be proved in Sect. IV. Recall (1.5).

Theorem 6 (Reduction of the Coulomb potential). *For any $0 < \lambda < 1$*

$$V_c(x_1, \dots, x_N; R_1, \dots, R_K) \geq - \sum_{i=1}^N W^\lambda(x_i) + \frac{1}{8} z^2 \sum_{j=1}^K D_j^{-1} \quad (2.4)$$

and, for x in the cell Γ_j , $W^\lambda(x) = W_j^\lambda(x) \equiv G_j^\lambda(x) + F_j^\lambda(x)$ with

$$G_j^\lambda(x) = z|x - R_j|^{-1} \quad (2.5)$$

$$F_j^\lambda(x) = \begin{cases} \frac{1}{2} D_j^{-1} (1 - D_j^{-2} |x - R_j|^2)^{-1} & \text{for } |x - R_j| \leq \lambda D_j \\ (1/\sqrt{2z} + \frac{1}{2}) |x - R_j|^{-1} & \text{for } |x - R_j| > \lambda D_j. \end{cases} \quad (2.6)$$

Theorem 6 says that when the electron-electron and nucleus-nucleus Coulomb repulsion is taken into account, V_c is bounded below by a *positive term* [the last term in (2.4)] consisting of a residue of the nucleus-nucleus repulsion (in fact one quarter of the nearest neighbor repulsion) and an attractive single particle part W^λ . In each cell Γ_j , W_j^λ is essentially the attraction to the nearest nucleus (this is the G_j part of W_j^λ); there is also a small attractive error F_j^λ .

There are two essential points in (2.4). One is that the charge z appearing in G_j is the same as in the original potential V_c . The other is the existence of the positive term. The error term F_j^λ can certainly be improved, especially the long-range part $|x - R_j| > \lambda D_j$; we have not tried to optimize F_j^λ .

It is interesting to compare our Theorem 6 with Baxter's Proposition 1 [1] which says that

$$V_c \geq -(1 + 2z) \sum_{j=1}^N \delta(x_j)^{-1} \quad (2.7)$$

with

$$\delta(x) = \min \{|x - R_j| \mid j = 1, \dots, K\} = |x - R_j| \quad \text{when } x \in \Gamma_j. \quad (2.8)$$

Fefferman and de la Llave [14] later improved this when $z = 1$ from $1 + 2z = 3$ to $8/3$. Our proof is completely different from both proofs of (2.7), as is Theorem 6

itself. To reiterate the essential points, our bound has the *correct* singularity near the nucleus (namely z and not $1+2z$) and it also has a *positive* repulsive term.

Step B. Control of the Coulomb Singularity in Balls

The following formula is well known. For $f \in L^2$ with Fourier transform $\hat{f}$,

$$(f, |p|f) = (2\pi)^{-3} \int |\hat{f}(p)|^2 |p| dp = (2\pi^2)^{-1} \iint |f(x) - f(y)|^2 |x - y|^{-4} dx dy. \quad (2.9)$$

One way to derive this formula is to write

$$(f, |p|f) = \lim_{t \downarrow 0} t^{-1} \{(f, f) - (f, e^{-t|p|}f)\}. \quad (2.10)$$

The convergence is a simple consequence of dominated convergence in Fourier space. The kernel of $\exp(-t|p|)$ can easily be calculated to be

$$e^{-t|p|}(x, y) = \pi^{-2} t [|x - y|^2 + t^2]^{-2}. \quad (2.11)$$

Inserting (2.11) in (2.10) yields (2.9).

A formula similar to (2.9) can be derived this way for $(p^2 + m^2)^{1/2}$ in place of $|p|$.

$$(f, (p^2 + m^2)^{1/2}f) = \frac{1}{4} \pi^{-2} m^2 \iint |f(x) - f(y)|^2 |x - y|^{-2} K_2(m|x - y|) dx dy, \quad (2.12)$$

where K_2 is a Bessel function. This follows from [11]

$$\exp[-t(p^2 + m^2)^{1/2}](x, y) = \frac{1}{2} \pi^{-2} m^2 t (|x - y|^2 + t^2)^{-1} K_2(m(|x - y|^2 + t^2)^{1/2}). \quad (2.13)$$

Starting with formula (2.9) we have

Theorem 7 (Kinetic energy in balls). *Let B be a ball of radius D centered at $z \in \mathbb{R}^3$ and let $f \in L^2(B)$. Define*

$$(f, |p|f)_B \equiv \frac{1}{2\pi^2} \iint_B |f(x) - f(y)|^2 |x - y|^{-4} dx dy \quad (2.14)$$

and assume this is finite. Then

$$(f, |p|f)_B \geq D^{-1} \int_B Q(|x - z|/D) |f(x)|^2 dx, \quad (2.15)$$

where $Q(r)$ is defined for $0 < r \leq 1$ by

$$\begin{aligned} Q(r) &= 2/(\pi r) - Y_1(r), \\ Y_1(r) &= \frac{2}{\pi(1+r)} + \frac{1+3r^2}{\pi(1+r^2)r} \ln(1+r) - \frac{1-r^2}{\pi(1+r^2)r} \ln(1-r) - \frac{4r}{\pi(1+r^2)} \ln r \\ &\leq 1.56712. \end{aligned} \quad (2.16)$$

The maximum of $Y_1(r)$ occurs at $r \approx 0.225975$ and was computed by S. Knabe. Note that $Y_1(|x|)$ is continuous for all $|x| \leq 1$.

Using (2.9) we have

Corollary. *If $B_1, \dots, B_K$ are disjoint balls in $\mathbb{R}^3$ centered at $R_1, \dots, R_K$ and with radii $D_1, \dots, D_K$,*

$$|p| \geq \frac{2}{\pi} \sum_{j=1}^K |x - R_j|^{-1} B_j(x) - \sum_{j=1}^K D_j^{-1} Y_1(|x - R_j|/D_j) B_j(x), \quad (2.17)$$

where $B_j(x)$ is the characteristic function of B_j .

Theorem 7 is proved in Sect. V. Theorem 12, which is the analogue of Theorem 7 with p^2 in place of $|p|$, is stated and proved in Sect. V.

Step C. Semiclassical Bounds and the Conclusion of the Proof of Theorem 1

The problem of showing that $H = \sum |p_i| + \alpha V_c \geq 0$ has been reduced to the following. In step A we showed that $H \geq \sum_1^N \tilde{h}_i + C$, where

$$\tilde{h}_i = |p_i| - \alpha W^\lambda(x_i), \quad (2.18)$$

$$C = \frac{1}{8} z^2 \alpha \sum_{j=1}^K D_j^{-1}. \quad (2.19)$$

If we write $|p| = \beta |p| + (1 - \beta) |p|$, with $\beta = z\alpha\pi/2$, then step B shows that it suffices to replace $\tilde{h}_i$ in H by h_i where

$$h_i = (1 - \beta) |p_i| - U(x_i), \quad (2.20)$$

$$U(x) = \alpha F_j^\lambda(x) + \beta D_j^{-1} Y_1(|x - R_j|/D_j) B_f(x) + z\alpha |x|^{-1} (1 - B_f(x)) \quad \text{when } x \in \Gamma_j. \quad (2.21)$$

Proving that $\sum_1^N h_i + C \geq 0$ for all numbers, N , of q -state fermions amounts to the following inequality in terms of density matrices satisfying $0 \leq \gamma \leq q$. [A density matrix is a positive definite trace class operator on $L^2(\mathbb{R}^3)$.]

$$\text{Tr} \gamma h \geq -C \quad \text{for all } \gamma, \quad (2.22)$$

with $h = (1 - \beta) |p| - U(x)$. ($\text{Tr} \gamma h$ is shorthand for $\sum_k (f_k, h f_k) \gamma_k$, where (f_k, γ_k) are the eigenfunctions and eigenvalues of γ .) For more details see [21].

The tool we shall use to prove (2.22) is Daubechies' extension of the Lieb-Thirring semiclassical bound from p^2 to $|p|$.

Theorem 8 (Daubechies). *Let γ be a density matrix satisfying $0 \leq \gamma \leq q$. (q need not be an integer.) Let $U(x)$ be any positive function in $L^4(\mathbb{R}^3)$. Then for $\mu > 0$,*

$$\text{Tr} \gamma (\mu |p| - U) \geq -0.0258 q \mu^{-3} \int U(x)^4 dx. \quad (2.23)$$

To complete the proof we merely insert (2.21) into (2.23). A simple bound is obtained by extending the integral over each Γ_j to an integral over all of $\mathbb{R}^3$. This will give K terms on the right side of (2.23) (each of which scales like D_j^{-1}) to be compared with the K terms in C (2.19). Our condition is then (recalling that $\beta = z\alpha\pi/2$)

$$\begin{aligned} & 0.0258 q (1 - \beta)^{-3} \left\{ \int_{|x| < 1} [\alpha F^\lambda(|x|) + \beta Y_1(|x|)]^4 dx \right. \\ & \left. + \int_{|x| > 1} [\alpha F^\lambda(|x|) + z\alpha |x|^{-1}]^4 dx \right\} \leq \frac{1}{8} z^2 \alpha \end{aligned} \quad (2.24)$$

for some choice of $0 < \lambda < 1$ and where

$$F^\lambda(r) = \begin{cases} \frac{1}{2} (1 - r^2)^{-1} & \text{for } 0 \leq r \leq \lambda \\ (\sqrt{2z} + \frac{1}{2}) r^{-1} & \text{for } \lambda \leq r. \end{cases} \quad (2.25)$$

The second integral ($|x| > 1$) in (2.24) (call it I_+) is easy to evaluate. It is independent of λ ,

$$I_+ = (4/\pi)^3 \beta^4 [1 + (2z)^{-1/2}]^8. \quad (2.26)$$

Next, the integral of Y_1^4 over $|x| < 1$ has been done numerically by S. Knabe. The following is actually an upper bound.

$$\int_{|x| < 1} Y_1(x)^4 dx = 7.6245 \equiv I_1. \quad (2.27)$$

We shall take $\lambda = 10/11$. Then

$$\int_{\lambda < |x| < 1} F^\lambda(x)^4 dx = (4\pi/10) \left[\frac{1}{2} + (2z)^{1/2} \right]^4 \equiv I_2, \quad (2.28)$$

$$\begin{aligned} \int_{|x| < \lambda} F^\lambda(x)^4 dx &\leq (\pi/4) \lambda \int_0^\lambda (1-r^2)^{-4} r dr \equiv I_3 \\ &= (\pi\lambda/24) [(1-\lambda^2)^{-3} - 1] = 22.645. \end{aligned} \quad (2.29)$$

To bound the first integral ($|x| < 1$) in (2.24) one can use the triangle inequality $\int (f+g+h)^4 \leq [\int (f^4)^{1/4} + (\int g^4)^{1/4} + (\int h^4)^{1/4}]^4$. Thus our condition for stability is satisfied if

$$\begin{aligned} 0.0258q(1-\beta)^{-3} \{ [1.6617\beta + 1.0588\alpha(\tfrac{1}{2} + (2z)^{1/2}) + 2.1815\alpha]^4 \\ + (4/\pi)^3 \beta^4 [1 + (2z)^{-1/2}]^8 \} \leq (2/\pi)^2 \beta^2 / 8\alpha. \end{aligned} \quad (2.30)$$

Let us rewrite the stability condition (2.30) as

$$q^{-1} z C(z) \geq \beta^3 (1-\beta)^{-3} \quad (2.31)$$

with $C(z)$ given by (1.8), namely

$$1/C(z) \equiv (0.0258) 4\pi \{ [1.6617 + 1.7258z^{-1} + 0.9533z^{-1/2}]^4 + (4/\pi)^3 [1 + (2z)^{-1/2}]^8 \}. \quad (2.32)$$

By taking the cube root in (2.31) we have that (2.31) is equivalent to the assertion that stability occurs if

$$\alpha \leq A_q(z) \equiv (2/\pi) z^{-1} \{ 1 + q^{1/3} z^{-1/3} C(z)^{-1/3} \}^{-1}. \quad (2.33)$$

Using the monotonicity in z for fixed α [8] mentioned in Sect. I, (2.33) can be improved to the statement that stability occurs if

$$\alpha \leq \sup \{ A_q(z') | z' \geq z \}, \quad (2.34)$$

and this is precisely Theorem 1. $\square$

Next, we address the question of finding a bound on α that depends only on β and not on z . For this purpose return to (2.31) and solve the equation $q^{-1} z C(z) = \beta^3 (1-\beta)^{-3}$. Since $z \rightarrow C(z)$ is monotone increasing, this equation has a unique solution. Call it $Z_q(\beta)$. Then stability occurs for any given β if

$$\alpha \leq \alpha_q^*(\beta) \equiv (2/\pi) \sup \{ \beta' / Z_q(\beta') | \beta' \geq \beta \}. \quad (2.35)$$

Again we have used the aforementioned fact that stability for (α, β') implies stability for (α, β) if $\beta \leq \beta'$.

Formula (2.35) is correct but lacks transparency. We shall now present a way to find a function $\alpha_q^{**}(\beta)$ which is less than or equal to $\alpha_q^*(\beta)$ but which has the same general features as $\alpha_q^*(\beta)$. It is this function, $\alpha_q^{**}(\beta)$ that is given in the corollary.

Choose an arbitrary z_0 . Let $\alpha_0 = A_q(z_0)$ and let $\beta_0 = (2/\pi)\alpha_0 z_0$. Define

$$\alpha_q^{**}(\beta) = \begin{cases} (2/\pi q)C(z_0)(1-\beta)^3\beta^{-2} & \text{if } \beta \geq \beta_0 \\ (2/\pi q)C(z_0)(1-\beta_0)^3\beta_0^{-2} & \text{if } \beta \leq \beta_0. \end{cases} \quad (2.36)$$

We claim that $\alpha \leq \alpha_q^{**}(\beta)$ implies stability. First, suppose that $\beta \geq \beta_0$. Then we have

$$z \equiv (2/\pi)\alpha^{-1}\beta \geq (2/\pi)[\alpha_q^{**}(\beta)]^{-1}\beta_0 \geq (2/\pi)[\alpha_q^{**}(\beta_0)]^{-1}\beta_0 \text{ (since } \beta \geq \beta_0) = z_0.$$

By the monotonicity of C , we have $C(z) \geq C(z_0)$. Therefore

$$z = (2/\pi)\alpha^{-1}\beta \geq (2/\pi)[\alpha_q^{**}(\beta)]^{-1}\beta \geq qC(z)^{-1}\beta^3(1-\beta)^{-3} \text{ (since } C(z) \geq C(z_0)).$$

This is (2.31).

Second, suppose that $\beta \leq \beta_0$. To prove the stability, we only have to verify (2.35). For this purpose, it suffices to show that $\alpha \leq \frac{2}{\pi}\beta_0/Z_q(\beta_0)$ with $Z_q(\beta_0)$ solving $q^{-1}zC(z) = \beta_0^3(1-\beta_0)^{-3}$. Since by definition $q^{-1}z_0C(z_0) = \beta_0^3(1-\beta_0)^{-3}$, we have from the uniqueness of the solution of the above equations that $Z_q(\beta_0) = z_0$ and $\alpha_q^{**}(\beta) = (2/\pi q)C(z_0)(1-\beta_0)^3\beta_0^{-2} = \frac{2}{\pi}\beta_0/Z_q(\beta_0)$. Hence $\alpha \leq \alpha_q^{**}(\beta)$ is the same as $\alpha \leq \frac{2}{\pi}\beta_0/Z_q(\beta_0)$ and thus stability occurs for (α, β) with $\alpha \leq \alpha_q^{**}(\beta)$ and $\beta \leq \beta_0$.

Let us choose $z_0 = 10$. Then $(2/\pi)C(10) = 0.062980$ and $\beta_0 = 0.49910$. This together with (2.36) proves the Corollary of Theorem 1. $\square$

III. Proof of Theorem 2 ($z\alpha \leq 2/\pi$)

In the proof of Theorem 1 we first reduced the many-body Coulomb potential to a one-body potential in Step A. Then we split the kinetic energy $|p|$ into two pieces. One of them was used to control the Coulomb singularity and the other was used to control the long range part of the potential. If the method of Theorem 1 is used when $z\alpha = 2/\pi$, all of $|p|$ must be used for the singularities and nothing remains to control the long-range potential. In this section both parts of the potential will be controlled without splitting $|p|$, but this requires inventing a suitable localization formula for $|p|$. We shall henceforth take $z\alpha = 2/\pi$; by the monotonicity in z , this case will cover all the cases $z\alpha \leq 2/\pi$.

There are five steps.

Step A is the same as before. The Coulomb potential V_c is replaced by a one-body potential W^λ plus a positive constant. Henceforth we shall take $\lambda = 0.97$ and omit the superscript on W .

Step B. Here we show that if $\chi_1(x)$ is a C^1 function which is approximately the characteristic function of a ball, and if γ is a density matrix with $0 \leq \gamma \leq q$ and if $\chi_2(x)$

is defined by $\chi_1(x)^2 + \chi_2(x)^2 = 1$, then

$$\begin{aligned} \text{Tr} \gamma(|p| - W) &\geq \text{Tr} \chi_1 \gamma \chi_1 (|p| - \text{potential energy correction} - W) \\ &\quad + \text{Tr} \chi_2 \gamma \chi_2 (|p| - \text{potential energy correction} - W) - q \cdot \text{const.} \end{aligned} \quad (3.1)$$

The important aspect of this inequality is this: It might have been thought that since $|p|$ is not a local operator, the potential energy corrections would have to be very long range. In fact they have support only inside a ball which is only slightly larger than the original ball. The long range nature of $|p|$ manifests itself in the term q -constant which depends on $\|\gamma\|$ but not on $N = \text{Tr} \gamma$.

Step C. The ball referred to in step B is taken to be B_1 centered at R_1 (see Sect. II). To control the first term on the right side of (3.1) we have to bound q times the sum of the negative eigenvalues of $|p| - \text{potential energy correction} - W$ in a ball, where W is the one-body potential defined in step A.

Step D. For the second term on the right side of (3.1), the localization process in steps B and C are repeated $K - 1$ times for nuclei, $2, \dots, K$. This finally leaves us with a term $\text{Tr} \chi_0 \gamma \chi_0 (|p| - \text{potential energy corrections})$ where χ_0 is essentially the characteristic function of the complement of the K balls. To estimate this term, Daubechies' semiclassical bound, Theorem 8, is used.

Step E. The above process leads to a lower bound on $\inf \text{spec}(H)$ in terms of certain integrals which depend on certain parameters that remain to be specified. These numerical facts are presented in this step. The details of the computation are given in Sect. VIII.

Step B. Localization of the Kinetic Energy

By way of comparison we begin by reminding the reader of the IMS localization formula (see [5, Theorem 3.2]) for $p^2 = -\Delta$ instead of $|p|$. Let $\chi_0, \chi_1, \dots, \chi_K$ be real valued functions on $\mathbb{R}^3$ satisfying

$$\sum_{j=0}^K \chi_j(x)^2 = 1 \quad \text{for all } x. \quad (3.2)$$

Then an elementary calculation yields the following operator identity.

$$-\Delta = \sum_{j=0}^K \chi_j(x) (-\Delta) \chi_j(x) - \sum_{j=0}^K |\nabla \chi_j(x)|^2. \quad (3.3)$$

This is a *localization of* $-\Delta$. If we assume additionally that χ_j has support in some set A_j (which are not pairwise disjoint, of course) then for any $f \in L^2(\mathbb{R}^3)$ and any arbitrary potential V ,

$$(f, [-\Delta + V(x)]f) = \sum_{j=0}^K (\chi_j f, [-\Delta + V(x) - U(x)] \chi_j f) \quad (3.4)$$

with

$$U(x) = \sum_{j=0}^K |\nabla \chi_j(x)|^2. \quad (3.5)$$

The advantage of (3.4) is that in the j^{th} term of (3.4) only $[V(x) - U(x)]1_{A_j}(x)$ appears [where $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ if $x \notin A$] and one can utilize different bounds on $V - U$ according to the region A_j under consideration. Furthermore, since $\chi_j f$ has support in A_j one can replace $-\Delta$ by the larger operator $-\Delta$ with Dirichlet boundary conditions on ∂A_j . The price one has to pay for all this is the negative potential operator $-U(x)$.

For the operator $|p|$ the following analogue of (3.3) is much more complicated because $|p|$ is not a local operator. We also state its generalization to $(p^2 + m^2)^{1/2}$. The proof is immediate starting with (2.9) and (2.12).

Theorem 9 (Localization of kinetic energy-general form). *Let $\chi_0, \dots, \chi_K$ be Lipschitz continuous functions satisfying (3.2). Then for any $f \in L^2(\mathbb{R}^3)$,*

$$(f, |p|f) = \sum_{j=0}^K (\chi_j f, |p| \chi_j f) - (f, Lf), \quad (3.6)$$

where L is a bounded operator with the kernel

$$L(x, y) = \frac{1}{2\pi^2} |x - y|^{-4} \sum_{j=0}^K [\chi_j(x) - \chi_j(y)]^2. \quad (3.7)$$

More generally,

$$(f, (p^2 + m^2)^{1/2} f) = \sum_{j=0}^K (\chi_j f, (p^2 + m^2) \chi_j f) - (f, L^{(m)} f), \quad (3.8)$$

where $L^{(m)}$ is a bounded operator with the kernel

$$L^{(m)}(x, y) = (2\pi)^{-2} m^2 |x - y|^{-2} K_2(m|x - y|) \sum_{j=0}^K [\chi_j(x) - \chi_j(y)]^2 \quad (3.9)$$

and K_2 is a Bessel function.

Formula (3.6) was proposed to us by M. Loss, to whom we are grateful.

A simple, but important corollary of Theorem 9 concerns q -state, density matrices. As defined in Sect. II, this is any bounded operator on $L^2(\mathbb{R}^3)$ which satisfies the operator inequality $0 \leq \gamma \leq q$ and for which $\text{Tr} \gamma < \infty$.

Corollary. *For any density matrix, γ ,*

$$\text{Tr} \gamma |p| = \sum_{j=0}^K \text{Tr} \gamma_j |p| - \text{Tr} \gamma L, \quad (3.10)$$

where $\gamma_j \equiv \chi_j \gamma \chi_j$, with χ_j being thought of as a multiplication operator.

To exploit (3.10) we now impose a condition on $\chi_0, \dots, \chi_K$. Let $R_1, \dots, R_K$ be distinct points in $\mathbb{R}^3$ (namely the nuclear coordinates) and let D_j be given by (2.3). The K disjoint balls $B_j = \{x | |x - R_j| < D_j\}$ were defined in Sect. II. Choose some $0 < \sigma < 1$ and consider the smaller balls

$$B_j^{(\sigma)} = \{x | |x - R_j| \leq (1 - \sigma) D_j\}. \quad (3.11)$$

Let $\chi_0, \dots, \chi_K$ satisfy (3.2) with χ_j supported in $B_j^{(\sigma)}$ for $j = 1, \dots, K$. The explicit choice for χ_j will be made in step D.

First, consider the case $K = 1$. We decompose the L of (3.7) into a long-range part, L^0 , and a short-range part, L_1^* , with $L = L^0 + L_1^*$. Furthermore, $L_1^*(x, y)$ vanishes if x or y is not in B_1 or if $|x - y| > \sigma$, namely

$$L_1^*(x, y) = \begin{cases} \pi^{-2} |x - y|^{-4} [1 - \chi_0(x)\chi_0(y) - \chi_1(x)\chi_1(y)] B_1(x)B_1(y) & \text{if } |x - y| \leq \sigma \\ 0 & \text{if } |x - y| > \sigma, \end{cases} \quad (3.12)$$

where $B_1(x) = 1$ if $x \in B_1$ and $B_1(x) = 0$ otherwise. Recall that $\chi_0(x)^2 + \chi_1(x)^2 = 1$ in the $K = 1$ case. With these conventions, we have the following theorem which will be proved in Sect. VI.

Theorem 10 (Localization of kinetic energy-explicit bound in the one-center case). *For $K = 1$, let L_1^* be given by (3.12) and $L^0 = L - L_1^*$, with L given by (3.7). For any positive function, h_1 , defined on the ball B_1 , let*

$$\theta_1(x) = h_1(x)^{-1} \int_{B_1} L_1^*(x, y) h_1(y) dy. \quad (3.13)$$

Let $\Omega_1 = \frac{1}{2} D_1^2 \text{Tr}(L^0)^2$, i.e.

$$\Omega_1 \equiv \frac{1}{2} D_1^2 \int [L(x, y) - L_1^*(x, y)]^2 dx dy \equiv I^{(1)} + I^{(2)}, \quad (3.14)$$

$$I^{(1)} = \frac{1}{2} \pi^{-4} D_1^2 \iint_{\substack{x, y \in B_1^{(\sigma)} \\ ||x| - |y|| \geq \sigma D_1}} |x - y|^{-8} [1 - \chi_0(x)\chi_0(y) - \chi_1(x)\chi_1(y)]^2 dx dy, \quad (3.15)$$

$$I^{(2)} = \pi^{-4} D_1^2 \iint_{\substack{x \in B_1^{(\sigma)} \\ y \notin B_1^{(\sigma)} \\ |y| - |x| \geq \sigma D_1}} |x - y|^{-8} [1 - \chi_0(x)]^2 dx dy. \quad (3.16)$$

Then, for any density matrix γ with $\|\gamma\| \leq q$, and any $\varepsilon > 0$,

$$\text{Tr} \gamma |p| \geq \text{Tr} \chi_1 \gamma \chi_1 (|p| - U_1^*(x)) + \text{Tr} \chi_0 \gamma \chi_0 (|p| - U_1^*(x)) - q(\varepsilon D_1)^{-1} \Omega_1, \quad (3.17)$$

where $U_1^*(x) = 0$ for $x \notin B_1$ and, for $x \in B_1$,

$$U_1^*(x) \equiv (\varepsilon/D_1) B_1^{(\sigma)}(x) + \theta_1(x). \quad (3.18)$$

Inequality (3.17) looks complicated, but it is not vastly different from (3.3). The first two terms in (3.17) are the localized kinetic energies (inside and outside the ball B_1). The U_1^* term is a potential energy correction like the U in (3.4), but this potential has support only in the ball B_1 . The last term is novel; it involves only the norm of γ and not a trace over γ . One might expect that the non-local nature of $|p|$ would give rise to a long range contribution to U , but these long range effects can be bounded by the norm of γ – as is done in the last term of (3.17).

Step C. Bound on Negative Eigenvalues in a Ball

Our goal is to give a lower bound to $\text{Tr} \chi_1 \gamma \chi_1 (|p| - W(x) - U_1^*(x))$. The following is our main tool. It will be proved in Sect. VII.

Theorem 11 (Lower bound to the short-range energy in a ball). *Let $C > 0$ and $R > 0$ and let*

$$H_{CR} = |p| - \frac{2}{\pi} |x|^{-1} - C/R \quad (3.19)$$

be defined on $L^2(\mathbb{R}^3)$ as a quadratic form. Let $0 \leq \gamma \leq q$ be a density matrix as before and let χ be any function with support in $B_R = \{x | |x| \leq R\}$. Then

$$\text{Tr } \bar{\chi} \gamma \chi H_{CR} \geq -4.4827 C^4 R^{-1} q \{ (3/4\pi R^3) \int |\chi(x)|^2 dx \}. \quad (3.20)$$

Remark. When $\chi \equiv 1$ in B_R then the factor in braces $\{ \}$ in (3.20) is 1.

To apply Theorem 11 to our case we take R in Theorem 11 to be $(1-\sigma)D_1$ and we take C to be an upper bound for $(1-\sigma)D_1 \{ \alpha W(x) + U_1^*(x) - (2/\pi)|x|^{-1} \} = (1-\sigma)D_1 \{ \alpha F_1(x) + U_1^*(x) \}$ in the ball $|x| \leq (1-\sigma)D_1$. This computation will be done in Step E.

Step D. The Negative Eigenvalues for the Long Range Potential

Associated with each ball B_j of radius D_j centered at R_j will be a cutoff function χ_j defined by

$$\chi_j(x) = \chi(|x - R_j|/D_j), \quad (3.21)$$

where the universal χ is given by

$$\chi(r) = \begin{cases} 1 & \text{for } r \leq 1-3\sigma \\ \cos[\pi(r-1+3\sigma)/4\sigma] & \text{for } 1-3\sigma \leq r \leq 1-\sigma \\ 0 & \text{for } 1-\sigma \leq r. \end{cases} \quad (3.22)$$

Here, it is important that $\sigma < 1/3$. We also choose a function $h_j(x)$ for $x \in B_j$,

$$h_j(x) = h(|x - R_j|/D_j), \quad (3.23)$$

$$rh(r) = \begin{cases} 1 & \text{for } r \leq 1-3\sigma \text{ and } 1-\sigma \leq r \leq 1 \\ 2\frac{1}{\pi}\sigma^{-1}|r-1+2\sigma| & \text{for } 1-3\sigma \leq r \leq 1-\sigma. \end{cases} \quad (3.24)$$

Starting with Theorem 10, Eq. (3.17), we choose some ε and compute $\Omega_1, \theta_1(x), U_1^*(x)$ using (3.13)–(3.16). We also compute some bound

$$C \geq (1-\sigma)D_1 \{ \alpha F_1(x) + U_1^*(x) \} \quad (3.25)$$

in $B_1^{(\sigma)}$. By scaling, C does not depend on D_1 . Then, using Theorem 11, Eq. (3.20), we have that

$$E = \inf_{\gamma} \text{Tr } \gamma(|p| - \alpha W) \geq -qA/D_1 + \inf_{\gamma} \text{Tr} (1 - \chi_1^2)^{1/2} \gamma (1 - \chi_1^2)^{1/2} (|p| - \alpha W - U_1^*). \quad (3.26)$$

The first term, qA/D_1 , is a sum of two pieces. One is the $q(\varepsilon D_1)^{-1} \Omega_1$ in (3.17); the other is the right side of (3.20) (call it qA_2). The sum is written as qA/D_1 because the various quantities that have been introduced scale in just the right way – so that A really is independent of D_1 and q .

For the second term on the right side of (3.26) we note the identity

$$(1 - \chi_1(x)^2) [\alpha W(x) + U_1^*(x)] = (1 - \chi_1(x)^2) [\alpha W(x) \beta_1(x) + U_1^*(x) \beta_1(x)], \quad (3.27)$$

where $\beta_1(x) = 1$ if $|x - R_1| \geq 1-3\sigma D_1$ and $\beta_1(x) = 0$ otherwise. Since $(1 - \chi_1^2)^{1/2} \gamma (1 - \chi_1^2)^{1/2}$ is a q -state density matrix whenever γ is, the last term in (3.26)

can be bounded below by

$$\inf_{\gamma} \text{Tr} \gamma (|p| - \alpha W(x) \beta_1(x) - U_1^*(x) \beta_1(x)). \quad (3.28)$$

Now we can apply Theorems 10 and 11 to (3.28), using the ball B_2 in place of B_1 . Since $U_1^*(x) = 0$ for $x \notin B_1$ we see that $(\alpha W(x) + U_1^*(x)) \beta_1(x) = \alpha W(x)$ for $x \notin B_1$. This process can be repeated until all the balls $B_1, \dots, B_K$ have been used. Our final result (with U_j^* defined as in (3.18) with R_1, D_1 replaced by R_j, D_j) is

$$E \geq -A \sum_{j=1}^K D_j^{-1} + \inf_{\gamma} \text{Tr} \gamma \left[|p| - \left(\alpha W + \sum_{j=1}^K U_j^*(x) \right) \prod_{j=1}^K \beta_j(x) \right]. \quad (3.29)$$

To bound the last term in (3.29) we use Theorem 8. This will result in a sum of K integrals, one for each cell Γ_j . As in the proof of Theorem 1, a further bound is obtained by pretending that each Γ_j extends to all of $\mathbb{R}^3$. Thus

$$E \geq -q(A+J) \sum_{j=1}^K D_j^{-1}, \quad (3.30)$$

where

$$J = 0.0258 \int_{|x| > 1-3\sigma} [(2/\pi)|x|^{-1} + \alpha F(|x|) + U^*(|x|)]^4 dx, \quad (3.31)$$

and where $F(r)$ is given in (2.25) with $\lambda = 0.97$, and $U^*(x)$ is given by (3.18) with $D_1 = 1$ there.

From (3.30) and (2.4), stability holds if

$$q(A+J) \leq \frac{1}{8} z^2 \alpha = (2\pi^2)^{-1} \alpha^{-1}. \quad (3.32)$$

Step E. Numerical Results

We take $\sigma = 0.3$ and $\varepsilon = 0.2077$ (recall that λ was previously chosen to be 0.97). Since all quantities have the correct length scaling, we shall refer everything to a standard ball of unit radius $D_1 = 1$. The following are the results of the computations given in Sect. VIII.

Starting with $\chi(r)$ in (3.22) we compute $\Omega_1 \equiv \Omega$ in (3.13)–(3.16),

$$\begin{aligned} I^{(1)} &= 0.05529, & I^{(2)} &= 0.06042, \\ \Omega &= I^{(1)} + I^{(2)} = 0.1157, & \varepsilon^{-1} \Omega &= 0.5571. \end{aligned} \quad (3.33)$$

From the definition (3.13) and (3.24) we find that $\theta_1(x) \equiv \theta(|x|)$ satisfies $\theta(r) \leq \theta^*(r)$ and

$$\theta^*(r) = \begin{cases} (3\pi/32)(2 - \sqrt{2})\sigma^{-1} = 0.5751 & \text{for } r \leq 1 - \sigma \\ (\pi/64)\sigma^{-5}(1 + 2\sigma - r)(1 - r)^3 & \text{for } 1 - \sigma < r \leq 1. \end{cases} \quad (3.34)$$

Using this we have, from (3.18), that

$$U^*(r) \leq \varepsilon B^{(\sigma)}(r) + \theta^*(r) \quad (3.35)$$

with $B^{(\sigma)}(r) = 1$ for $r < 1 - \sigma = 0.7$ and $B^{(\sigma)}(r) = 0$ otherwise.

Next, we want to find some C satisfying (3.25). Since $\lambda = 0.97 > 1 - \sigma = 0.7$, we need only concern ourselves with the first line of (2.25). Note that α appears in (3.25)

in the form $\alpha F_1(x)$ and, since $F_1(x)$ does not depend on z in the region $r < \lambda$, the quantity $\alpha F_1(x)$ is proportional to α when $z\alpha$ is fixed. Our goal is to prove stability when $\alpha < 1/47q \leq 1/47$, and therefore we can replace $\alpha F_1(x)$ by $F_1(x)/47$ in (3.25). Then

$$C = 0.7 \{0.02086 + 0.2077 + 0.5751\} = 0.5629 \quad (3.36)$$

satisfies (3.25) for $r < 1 - \sigma = 0.7$.

The right side of (3.20) (with $R = 1 - \sigma = 0.7$) can now be easily calculated. It is

$$qA_2 = 0.1661q. \quad (3.37)$$

Adding $\varepsilon^{-1}\Omega$ and A_2 we have

$$A = 0.7232. \quad (3.38)$$

Finally, the integral in (3.31) must be computed. To bound $\alpha F(r)$ we can use $(1/47)F(r)$ for $r < \lambda$, while for $r > \lambda$ we write $z = 2/\pi\alpha$ in (2.25). When $r > \lambda$ this results in two terms in αF , one of which is proportional to $\alpha^{1/2}$ and the other to α . In both terms we can take $\alpha = 1/47$. Thus, we bound $\alpha F(r)$ by $0.1753/r$ for $r > \lambda$ and by $(1/94)(1 - r^2)^{-1}$ for $r < \lambda$. We then find that

$$\begin{aligned} (0.0258)^{-1} (4\pi)^{-1} J \leq & \int_{0.1}^{0.7} [2/\pi r + (1/94)(1 - r^2)^{-1} + 0.2077 + 0.5751]^4 r^2 dr \\ & + \int_{0.7}^{0.97} [(2/\pi r + (1/94)(1 - r^2)^{-1} + 20.20(1.6 - r)(1 - r)^3]^4 r^2 dr \\ & + \int_{0.97}^1 [2/\pi r + 0.1753/r + 20.20(1.6 - r)(1 - r)^3]^4 r^2 dr \\ & + \int_1^\infty [2/\pi r + 0.1753r^{-1}]^4 r^2 dr. \end{aligned} \quad (3.39)$$

The first integral, J_1 , can be bounded by replacing $(1 - r^2)^{-1}$ by $(1 - (0.7)^2)^{-1}$ and then doing the integral analytically. The second integral, J_2 , was done on a computer. In the third integral, J_3 , $1.6 - r$ was replaced by $1.6 - 0.97$ and $(1 - r^3)$ was replaced by $(1 - 0.97)^3$; it was then done analytically. The fourth integral, J_4 , can be done analytically. We find $J_1 \leq 4.435$, $J_2 \leq 0.17$, $J_3 \leq 0.0135$, and $J_4 \leq 0.435$. Thus

$$J \leq 1.64 \quad (3.40)$$

and, from (3.32), stability occurs if $\alpha q < 1/47$. This completes the proof of Theorem 2. $\square$

IV. An Electrostatic Inequality

Our goal here is to prove Theorem 6 about the Coulomb potential V_c given in (1.5). A similar theorem can be derived for the Yukawa potential $|x|^{-1} \exp(-\mu|x|)$, but we shall not do so here. We recall the definition (2.2) of the K Voronoi cells $\Gamma_1, \dots, \Gamma_K$ for K nuclei located at $R_1, \dots, R_K \in \mathbb{R}^3$, and also the radii D_j in (2.3) which is the distance of R_j to $\partial\Gamma_j$. Since Theorem 6 is trivial when $K = 1$, we shall assume henceforth that $K > 1$. We set

$$V(x) = \sum_{j=1}^K |x - R_j|^{-1}, \quad (4.1)$$

which is the potential of K nuclei of unit charge located at the R_j , and

$$\delta(x) = \min \{|x - R_j| \mid 1 \leq j \leq K\}, \quad (4.2)$$

which is the distance of a particle at x to the set of K nuclei. We set

$$\Phi(x) = V(x) - \delta(x)^{-1}, \quad (4.3)$$

which is the potential of all the nuclei except for the nucleus in the cell I_j in which x is located. Φ is continuous but not differentiable.

Let ν be any Borel measure (possibly signed) on $\mathbb{R}^3$. We say that ν is a bounded measure if $|\nu|(\mathbb{R}^3) < \infty$. In this case $\int \Phi(x) d\nu(x)$ is well defined since Φ is continuous and bounded. We define

$$\mathcal{E}_{\Phi, z}(\nu) = \frac{1}{2} \iint |x - y|^{-1} d\nu(x) d\nu(y) - z \int \Phi(x) d\nu(x) + z^2 \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1}. \quad (4.4)$$

The first term on the right side of (4.4) is well defined (in the sense that it is either finite or $+\infty$) since $|x - y|^{-1}$ is a positive definite kernel. The following is basic to our analysis.

Lemma 1. *Let ν be any bounded measure, let $z > 0$ and let Φ be given by (4.3). Then*

$$\mathcal{E}_{\Phi, z}(\nu) \geq \frac{1}{8} z^2 \sum_{j=1}^K D_j^{-1}. \quad (4.5)$$

Proof. There is a (positive) measure μ that satisfies the equation

$$|x|^{-1} * \mu = z\Phi \quad (4.6)$$

and μ has support on $\partial\Gamma \equiv \bigcup_{j=1}^K \partial I_j$. In fact, μ can be computed explicitly as

$$\mu = -(z/4\pi) \Delta \Phi. \quad (4.7)$$

More precisely, $\partial\Gamma$ consists of pieces of 2 dimensional planes separating some I_i from some I_j ; on ∂I_j

$$d\mu(x) = -(z/2\pi) \mathbf{n} \cdot \nabla |x - R_j|^{-1} d^2x, \quad (4.8)$$

where d^2x is two-dimensional Lebesgue measure on ∂I_j , and $\mathbf{n}$ is the unit normal pointing out of I_j . Let

$$A = -\frac{1}{2} z \int \delta(x)^{-1} d\mu(x). \quad (4.9)$$

Then

$$\begin{aligned} \frac{1}{2} \iint |x - y|^{-1} d\mu(x) d\mu(y) &= \frac{1}{2} \int \Phi(x) d\mu(x) = \frac{z}{2} \sum_{j=1}^K \int |x - R_j|^{-1} d\mu(x) + A \\ &= \frac{z^2}{2} \sum_{j=1}^K \Phi(R_j) + A = z^2 \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1} + A. \end{aligned} \quad (4.10)$$

On the other hand, if each part of $\partial\Gamma$ is counted twice we obtain

$$A = (z^2/8\pi) \sum_{j=1}^K \int_{\partial I_j} |x - R_j|^{-1} \mathbf{n} \cdot \nabla |x - R_j|^{-1} d^2x. \quad (4.11)$$

Let I_j denote the integral in (4.11). The integrand is $\frac{1}{2} \mathbf{n} \cdot \nabla |x - R_j|^{-2}$. With A_j denoting the complement of Γ_j in $\mathbb{R}^3$ (so that $\partial A_j = \partial \Gamma_j$) we have

$$I_j = \frac{1}{2} \int_{\partial \Gamma_j} \mathbf{n} \cdot \nabla |x - R_j|^{-2} d^2x = -\frac{1}{2} \int_{A_j} \Delta |x - R_j|^{-2} dx = -\int_{A_j} |x - R_j|^{-4} dx. \quad (4.12)$$

For convenience in evaluating (4.12) we can take $R_j = 0$ and assume that A_j contains the half-space $\{(x, y, z) | x \geq D_j\}$; the reason for this is that (assuming $D_j \neq \infty$) there is another nucleus at some R_i such that the midplane between R_j and R_i is given (after rotation of coordinates) by $\{(x, y, z) | x = D_j\}$. Thus

$$I_j \leq -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dy dz \int_{D_j}^{\infty} dx (x^2 + y^2 + z^2)^{-2} = -\pi/D_j, \quad (4.13)$$

and therefore

$$A \leq -\frac{1}{8} z^2 \sum_{j=1}^K D_j^{-1}. \quad (4.14)$$

Using (4.6) and (4.10) we have that

$$\mathcal{E}_{\Phi, z}(v) = \frac{1}{2} \iint |x - y|^{-1} d(v - \mu)(x) d(v - \mu)(y) - A. \quad (4.15)$$

The integral in (4.15) is nonnegative (since $|x - y|^{-1}$ is positive definite), and the lemma follows from (4.14). $\square$

Proof of Theorem 6. There are N points $x_1, \dots, x_N$. If x_i is in some cell Γ_j we shall replace the unit point charge at x_i by a unit charge distributed on a sphere S_i but, in general, the center of S_i will not be x_i and the charge distribution on S_i will not be uniform. Also, S_i is not always contained entirely in Γ_j . (If x_i is in more than one Γ_j then an arbitrary choice can be made.) The definition of S_i and the charge distribution v_i on S_i is the following:

(i) If $|x_i - R_j| \leq \lambda D_j$, then S_i is the sphere $\partial B_j = \{x | |x - R_j| = D_j\}$. The charge v_i is determined so that its (continuous) potential $V_i \equiv |x|^{-1} * v_i$ satisfies

$$V_i(x) = \begin{cases} |x - x_i|^{-1} & \text{for } |x - R_j| \geq D_j \\ |x - x_i^*|^{-1} |x_i - R_j|^{-1} D_j & \text{for } |x - R_j| \leq D_j, \end{cases} \quad (4.16)$$

where x_i^* is the image of x_i with respect to S_j , namely

$$x_i^* - R_j = D_j^2 |x_i - R_j|^{-2} (x_i - R_j). \quad (4.17)$$

The potential $V_i(x)$ is harmonic inside and outside B_j , and v_i can be computed from the formula $-\Delta V_i = 4\pi v_i$, but we shall not need this. It is important to note that v_i is nonnegative.

(ii) If $|x_i - R_j| > \lambda D_j$ and $x_i \in \Gamma_j$, then S_i is a sphere centered at x_i and of radius t_i given by

$$t_i = |x_i - R_j| (1 + \sqrt{2z})^{-1}. \quad (4.18)$$

The charge distribution v_i on S_i is the uniform one with unit total charge.

Now we apply Lemma 1 with

$$v = \sum_{i=1}^N v_i. \quad (4.19)$$

In order to utilize inequality (4.5) it is necessary to relate $\mathcal{E}_{\Phi,z}(v)$ to V_c . The last term in (4.4) is, of course, exactly the nuclear repulsion. The first term on the right side of (4.4) (call it I) satisfies

$$I = \sum_{1 \leq i < k \leq N} \int \int |x-y|^{-1} dv_i(x) dv_k(y) + \frac{1}{2} \sum_{i=1}^N \int \int |x-y|^{-1} dv_i(x) dv_i(y). \quad (4.20)$$

Each $v_i v_k$ integral in (4.20) is less than or equal to $|x_i - x_k|^{-1}$. This is so because, by construction

$$(|x|^{-1} * v_i)(x) \leq |x - x_i|^{-1}, \quad \text{all } x, \quad (4.21)$$

and hence

$$\int (|x|^{-1} * v_i)(x) dv_k(x) \leq (|x|^{-1} * v_k)(x_i) \leq |x_k - x_i|^{-1}. \quad (4.22)$$

The $v_i v_i$ integral in (4.20) is just the self energy of v_i . Call it e_i . There are two cases.

(i) $|x_i - R_j| \leq \lambda D_j$. Then, from (4.16)

$$\begin{aligned} e_i &= \int \int |x-y|^{-1} dv_i(x) dv_i(y) = \int |x - x_i|^{-1} dv_i(x) = V_i(x_i) \\ &= |x_i - x_i^*|^{-1} |x_i - R_j|^{-1} D_j = D_j^{-1} (1 - D_j^{-2} |x_i - R_j|^2)^{-1}. \end{aligned} \quad (4.23)$$

(ii) $|x_i - R_j| > \lambda D_j$ and $x_i \in \Gamma_j$. Here $e_i = 1/t_i$ since v_i is uniformly distributed on a sphere of radius t_i .

To summarize,

$$I \leq \sum_{1 \leq i < k \leq N} |x_i - x_k|^{-1} + \frac{1}{2} \sum_{i=1}^N \left\{ \begin{array}{ll} \text{Eq. (4.23)} & \text{in case (i)} \\ 1/t_i & \text{in case (ii)} \end{array} \right\}. \quad (4.24)$$

The second term on the right side of (4.4) is a sum of $z \int \Phi dv_i$. Again, there are two cases.

(i) $|x_i - R_j| \leq \lambda D_j$. From the definition of W and the fact that $(|x|^{-1} * v_i)(x) = |x - x_i|^{-1}$ when $x \notin \Gamma_j$, we have

$$\int \Phi(x) dv_i(x) = \sum_{k=1}^K |x_i - R_k|^{-1} - |x_i - R_j|^{-1}. \quad (4.25)$$

(ii) $|x_i - R_j| > \lambda D_j$ and $x_i \in \Gamma_j$. By the definition of Φ

$$\int \Phi(x) dv_i(x) = \sum_{k=1}^K \int |x - R_k|^{-1} dv_i(x) - \int \delta(x)^{-1} dv_i(x), \quad (4.26)$$

where $\delta(x)$ is the distance to the nearest nucleus. Since every R_k (including R_j) is outside S_i , the first term in (4.26) is merely $\sum_{k=1}^K |x_i - R_k|^{-1}$. The difficulty in estimating the second term in (4.4) stems from the fact that v_i can have support in several cells – not just Γ_j . We have, however, that for $|x - x_i| = t_i$ and any k ,

$$|x - R_k| + t_i = |x - R_k| + |x - x_i| \geq |R_k - x_i| \geq |R_j - x_i|. \quad (4.27)$$

Hence $\delta(x) \geq |R_j - x_i| - t_i$, and therefore in case (ii),

$$\int \Phi(x) dv_i(x) \geq \sum_{k=1}^K |x_i - R_k|^{-1} - (|R_j - x_i| - t_i)^{-1}. \quad (4.28)$$

Using these inequalities and the definition (4.18) we find that

$$\mathcal{E}_{\Phi, z}(v) \leq V_c + \sum_{i=1}^N W^\lambda(x_i), \quad (4.29)$$

with $W^\lambda(x)$ given in (2.5), (2.6). This, together with Lemma 1, proves Theorem 6. $\square$

V. Simple Localization of the Kinetic Energy

Here we shall prove Theorem 7, but before doing so let us motivate Theorem 7 by stating the analogous Theorem 12 below for p^2 instead of $|p|$. This latter theorem is simple to prove, but we have not seen it in the literature.

Theorem 12 (The energy of p^2 in balls). *Let B be a ball of radius R centered at $z \in \mathbb{R}^3$ and let $f \in L^2(B)$ and $\nabla f \in L^2(B)$. Define*

$$(f, p^2 f)_B = \int_B |\nabla f(x)|^2 dx. \quad (5.1)$$

Then

$$(f, p^2 f)_B \geq R^{-2} \int_B H((x-z)/R) |f(x)|^2 dx, \quad (5.2)$$

where $H(x)$, for $|x| < 1$, is any function of the form $H(x) = -h^{-1}(x)\Delta h(x)$ and where h is a smooth, strictly positive function with vanishing normal derivative on the boundary $|x| = 1$. In particular, by taking $h(x) = (|x|^2 + t)^{-1/4} \exp[\frac{1}{4}|x|^2/(1+t)]$, and then letting $t \rightarrow 0$ (using Fatou's lemma) we have that (5.2) holds with

$$H(x) = \frac{1}{4}|x|^{-2} - Y_2(|x|), \quad Y_2(r) = 1 + \frac{1}{4}r^2. \quad (5.3)$$

Remark. It is important to note that $\frac{1}{4}$, the coefficient of the $|x|^{-2}$ singularity, is precisely the sharp constant for the uncertainty principle in all of $\mathbb{R}^3$, $(f, p^2 f) \geq \frac{1}{4} \int |f|^2 |x|^{-2} dx$.

Proof. Write $f(x) = g(x)h(x)$ so that $\nabla f = h\nabla g + g\nabla h$. Then

$$\int_B |\nabla f|^2 = \int_B h^2 |\nabla g|^2 + \int_B |g|^2 (\nabla h)^2 + \int_B (\nabla g^2) h \nabla h. \quad (5.4)$$

Integrating the last integral by parts

$$\int_B |\nabla f|^2 \geq - \int_B g^2 h \Delta h = \int_B f^2 H. \quad (5.5)$$

Equation (5.3) is merely a calculation. $\square$

We turn now to the problem of proving Theorem 7 which is the analogue of Theorem 12 for

$$(f, |p| f)_B = (2\pi^2)^{-1} \int_B \int_B |f(x) - f(y)|^2 |x - y|^{-4} dx dy. \quad (5.6)$$

If B is $\mathbb{R}^3$ then this is just $(f, |p| f)$; see (2.9).

Proof of Theorem 7. Without loss of generality we can take $z = 0$ and $R = 1$. First, we regularize $|x - y|^{-4}$ to $L_t(x, y) = (|x - y|^2 + t)^{-2}$. The theorem will follow by letting $t \rightarrow 0$ and using dominated convergence and Fatou's lemma.

With L_t in place of $|x-y|^{-4}$ we have

$$(f, |p| f)_{B,t} = \pi^{-2} \int_B |f(x)|^2 K_t(x) dx - \pi^{-2} \int_B \int_B f(x) \bar{f}(y) L_t(x, y) dx dy, \quad (5.7)$$

$$K_t(x) = \int_B L_t(x, y) dy. \quad (5.8)$$

The second integral in (5.7) can be bounded above using the Schwarz inequality as follows. Choose a real valued function h with $h(x) > 0$ for all $|x| \leq 1$. Then

$$\begin{aligned} \int \bar{f} \bar{f} L_t &= \int_B \int_B [f(x)h(y)^{1/2}/h(x)^{1/2}] [\bar{f}(y)h(x)^{1/2}/h(y)^{1/2}] L_t(x, y) dx dy \\ &\leq \int_B |f(x)|^2 \eta_t(x) dx \end{aligned} \quad (5.9)$$

with

$$\eta_t(x) = h(x)^{-1} \int_B L_t(x, y) h(y) dy. \quad (5.10)$$

We make the choice that h is radial, i.e. $h(x) = h(r)$ with $r = |x|$. To compute K_t and η_t we can do the angular y integration. With $|y| = s$ we have

$$l_t(r, s) \equiv \int L_t(x, y) d\omega_y = [\pi/rs] \{[(r-s)^2 + t]^{-1} - [(r+s)^2 + t]^{-1}\}. \quad (5.11)$$

Combining (5.7)–(5.11) we have that

$$(f, |p| f)_{B,t} \geq \int_B |f(x)|^2 Q_t(|x|) dx, \quad (5.12)$$

with

$$Q_t(r) = \pi^{-2} \int_0^1 l_t(r, s) [1 - h(s)/h(r)] s^2 ds. \quad (5.13)$$

Finally, we choose

$$h(r) = (1 + r^2)/r. \quad (5.14)$$

(Note that $dh/dr = 0$ at $r = 1$.) The integrand in (5.13) is then

$$\pi r^{-1} (1 + r^2)^{-1} (s - r) (1 - rs) \{[(r-s)^2 + t]^{-1} - [(r+s)^2 + t]^{-1}\}. \quad (5.15)$$

At this point we can let $t \rightarrow 0$ by recognizing that the integral in (5.13) becomes a principal value integral in the limit, i.e. $Q_t \rightarrow Q$ with

$$Q(r) = 4\pi^{-1} (1 + r^2)^{-1} \oint (s - r)^{-1} (r + s)^{-2} (s - rs^2) ds. \quad (5.16)$$

To do this integral (call it I) we set

$$I_1 = \oint_0^1 (s - r)^{-1} (r + s)^{-2} s ds = [2r(1 + r)]^{-1} - (4r)^{-1} \ln[(1 + r)/(1 - r)]. \quad (5.17)$$

The remainder of I (namely the rs^2 term) is

$$- \int_0^1 rs(r + s)^{-2} ds - r^2 I_1 = -r \ln[(1 + r)/r] + r(r + 1)^{-1} - r^2 I_1. \quad (5.18)$$

By combining (5.17), (5.18), Eq. (2.16) is derived. The maximum of $Y_1(r)$ was computed numerically by S. Knabe. $\square$

With the help of Theorems 7 and 12, the proof of Theorem 5, which was stated in Sect. 1, can now be given.

Proof of Theorem 5. Fix $0 < L < N$ and $M = N - L$ and consider any partition $P = (\pi_1, \pi_2)$ of $\{1, \dots, N\}$ into two disjoint sets with L integers in π_1 and M integers in π_2 . There are $\binom{N}{L}$ such partitions. For each P we define

$$\delta_i(\pi_2) = \min \{|x_i - x_j| \mid j \in \pi_2 \text{ and } j \neq i \text{ if } i \in \pi_2\}. \quad (5.19)$$

First the operator $|p|$ will be considered. Define the N -particle operator

$$h_P = \sum_{i \in \pi_1} |p_i| - \lambda \sum_{i \in \pi_1} \delta_i(\pi_2)^{-1} + \alpha \sum_{i \in \pi_2} \delta_i(\pi_2)^{-1} \quad (5.20)$$

for some $\lambda, \alpha > 0$ to be determined later. Let the N -particle operators H and $\hat{H}$ be given by

$$H = \binom{N}{L}^{-1} \frac{N}{L} \sum_P h_P, \quad (5.21)$$

$$\hat{H} = \sum_{i=1}^N |p_i| - C_1 q^{-1/3} \sum_{i=1}^N \delta_i^{-1}. \quad (5.22)$$

If H and $\hat{H}$ are compared we observe that the $|p_i|$ terms are identical. The potential energy terms are more complicated, but we wish to choose λ and α so that $\hat{H} \geq H$. To this end, fix $x_1, \dots, x_N$ and let $x_{j(i)}$ be a nearest neighbor of x_i , that is $|x_{j(i)} - x_i| = \min \{|x_k - x_i| \mid k \neq i\}$. It is obvious that $\delta_i(\pi_2)^{-1} \leq \delta_i^{-1}$, so that the last term in (5.20), when summed on P , is at most $\alpha \tau \sum_1^N \delta_i^{-1}$, where

$$\tau = \binom{N}{L}^{-1} \frac{N}{L} \binom{N-1}{L} = \frac{N-L}{L}. \quad (5.23)$$

To bound the middle, or λ , term in h_P we note that for each $i \in \{1, \dots, N\}$ there will be $\binom{N-2}{L-1}$ partitions in which $i \in \pi_1$ and $j(i) \in \pi_2$. Therefore this middle sum in h_P , when summed on all partitions, is at least $\lambda v \sum_1^N \delta_i^{-1}$, where

$$v = \binom{N}{L}^{-1} \frac{N}{L} \binom{N-2}{L-1} = \frac{N-L}{N-1}. \quad (5.24)$$

Consequently, $\hat{H} \geq H$ if

$$C_1 q^{-1/3} \leq (N-L) [\lambda(N-1)^{-1} - \alpha L^{-1}]. \quad (5.25)$$

Assuming (5.25), Theorem 5 will be proved if we show that $(\psi, h_P \psi) \geq 0$ for every P . Since permutation of the labels in π_1 and π_2 is irrelevant, it suffices to prove this for any one P . To this end we henceforth change notation so that $x_1, \dots, x_L \in \mathbb{R}^3$ are the variables in the π_1 block and $R_1, \dots, R_M \in \mathbb{R}^3$ are the variables in the π_2 block. Obviously we can assume that the R_i are fixed and distinct and that ψ is then a function of $x_1, \dots, x_L$ with q -state Fermi statistics. We

shall also drop the subscript P on h_P . Thus, we want to show that $h \geq 0$ for all choices of the R_i . Since h is a sum of one-body operators, we have to show that for any density matrix γ with $0 \leq \gamma \leq q$,

$$\text{Tr} \gamma(|p| - V) \geq -\alpha \sum_{j=1}^M (2D_j)^{-1}, \quad (5.26)$$

where $V(x)$ and D_j are defined by

$$V(x) = -\lambda \delta(x)^{-1}, \quad (5.27)$$

$$2D_j = \min \{|R_j - R_k| | k = 1, \dots, M \text{ but } k \neq j\}, \quad (5.28)$$

$$\delta(x) = \min \{|x - R_j| | j = 1, \dots, M\}. \quad (5.29)$$

Under the assumption that $\lambda < 2/\pi$, we write $|p|$ as the sum of two pieces $|p| = (\lambda\pi/2)|p| + (1 - \lambda\pi/2)|p|$. We also introduce the Voronoi cells $\Gamma_j = \{x | |x - R_j| \leq |x - R_k| \text{ for all } k \neq j\}$ and the balls $B_j \subset \Gamma_j$ defined by $B_j = \{x \in \Gamma_j | |x - R_j| \leq D_j\}$. Obviously

$$(f, |p|f) \geq \sum_{j=1}^M (f, |p|f)_{B_j}, \quad (5.30)$$

where the right side is the sum of the kinetic energies in the balls B_j defined in Theorem 7, (2.14). Using Theorem 7, we have that

$$(\lambda\pi/2)(f, |p|f) \geq (\lambda\pi/2) \sum_{j=1}^M D_j^{-1} \int_{B_j} |f(x)|^2 Q(|x - R_j|/D_j) dx, \quad (5.31)$$

with Q given by (2.16). Hence

$$\text{Tr} \gamma(|p| - V) \geq \text{Tr} \gamma[(1 - \lambda\pi/2)|p| - \lambda W], \quad (5.32)$$

where W is given in each Γ_j by

$$W_j(x) = \begin{cases} |x - R_j|^{-1} & \text{if } |x - R_j| > D_j \\ (\pi/2)D_j^{-1} Y_1(|x - R_j|/D_j) & \text{if } |x - R_j| \leq D_j \end{cases} \quad (5.33)$$

with Y_1 given in (2.16).

Next, we use the Daubechies bound, Theorem 8,

$$\text{Tr} \gamma[(1 - \lambda\pi/2)|p| - \lambda W] \geq -0.0258q[1 - \lambda\pi/2]^{-3} \lambda^4 \int W(x)^4 dx. \quad (5.34)$$

The integral in (5.34) is a sum of integrals over each Γ_j . To obtain a bound we shall merely integrate each $|x - R_j|$ term in W [see (5.33)] over all $|x - R_j| > D_j$ and omit the restriction that $x \in \Gamma_j$. The integral outside each ball B_j is thus

$$\int_{\sim B_j} W_j^4 = 4\pi/D_j. \quad (5.35)$$

The integral inside B_j is (see (2.27))

$$\int_{B_j} W_j^4 = (\pi/2)^4 D_j^{-1} \int_{|x| < 1} Y_1(x)^4 dx = 46.418/D_j. \quad (5.36)$$

Combining (5.34)–(5.36) we find that (5.26) is satisfied provided

$$qA\lambda^4(1 - \lambda\pi/2)^{-3} \leq \frac{1}{2}\alpha \quad (5.37)$$

with

$$A = 0.0258[4\pi + 46.418] = 1.522 \quad (5.38)$$

and provided $\lambda < 2/\pi$. We shall choose α so that (5.37) is an equality. We shall also write $\lambda = Xq^{-1/3}$. Then (5.25) is satisfied if C_1 satisfies the following for some $0 \leq X \leq 2/\pi$ and some $0 < L < N$:

$$C_1 \leq (N-L)[X(N-1)^{-1} - AX^4(1-X\pi/2)^{-3}L^{-1}]. \quad (5.39)$$

(Here we have used the bound that $\lambda\pi/2 < X\pi/2$, which holds since $q \geq 1$.)

Consider the case $N \geq 3$. To utilize (5.39) we make the following choices

$$X = 1/5 \quad \text{and} \quad L = \{(B/X)^{1/2}N\}, \quad (5.40)$$

where $B \equiv AX^4(1-X\pi/2)^{-3} = 0.0075486$ and where $\{a\}$ denotes the smallest integer $\geq a$. Write $L = l + \varepsilon$ with $l = N(B/X)^{1/2}$ and $0 \leq \varepsilon < 1$. We claim that when $N \geq 3$,

$$(L-1)X/(N-1) + BN/L \leq lX/N + BN/l. \quad (5.41)$$

Assuming this for the moment, we would then have that (5.39) is satisfied with

$$C_1 = (X^{1/2} - B^{1/2})^2 \geq 0.129, \quad (5.42)$$

which proves Theorem 5 when $N \geq 3$. If $N = 1$ there is nothing to prove. If $N = 2$, Theorem 3 is trivial because it asserts that

$$|p_1| + |p_2| \geq 0.129q^{-1/3}|x_1 - x_2|^{-1}, \quad (5.43)$$

but we already have the simple bound $|p_1| \geq (2/\pi)|x_1 - x_2|^{-1}$ for all x_2 .

To prove (5.41), insert $L = l + \varepsilon$ in the left side and multiply by $N(N-1)L$ (recalling that $l \equiv N(B/X)^{1/2}$). Then (5.41) is equivalent to

$$Nl - l(l + 2\varepsilon) + N\varepsilon(1 - \varepsilon) \geq 0. \quad (5.44)$$

Since $l < N/5$, (5.44) holds for $N \geq 3$.

The proof for p^2 in place of $|p|$ follows the same route, but using Theorem 12 in place of Theorem 7 and using the Lieb-Thirring [25] bound in place of the Daubechies bound. This is

$$\text{Tr}(\mu p^2 - \lambda W) \geq -q\sigma\mu^{-3/2}\lambda^{5/2} \int W(x)^{5/2} dx.$$

The best bound for σ is obtained in [22] and is $\sigma = 0.040305$. We split the operator p^2 into $4\lambda p^2 + (1 - 4\lambda)p^2$, and take the μ above to be $(1 - 4\lambda)$. Using Theorem 12, W is given in each cell I_j by

$$W(x) = \begin{cases} |x - R_j|^{-2} & \text{if } |x - R_j| > D_j \\ 4D_j^{-2}Y_2(|x - R_j|/D_j) & \text{if } |x - R_j| < D_j. \end{cases} \quad (5.45)$$

The analogue of (5.35), (5.36) using $Y_2(r) = 1 + r^2/4$, is

$$w \equiv D_j^2 \int_{\mathbb{R}^3} W_j(x)^{5/2} dx = 2\pi + 128\pi \int_0^1 (1 + r^2/4)^{5/2} r^2 dr.$$

Using $(1 + r^2/4)^{1/2} \leq 1 + r^2/8$ in the above integral we find $w < 198.2$.

Setting $\lambda = Xq^{-2/3}$, the analogue of (5.39) is

$$C_2 \leq (N-L)[X(N-1)^{-1} - AX^{5/2}(1-4X)^{-3/2}L^{-1}] \quad (5.46)$$

with $A = \sigma w = 7.988$. For $N \geq 3$ we make the following choices:

$$X = 1/20 \quad \text{and} \quad L = \{(B/X)^{1/2}N\}, \quad (5.47)$$

with $B = AX^{5/2}(1-4X)^{-3/2} = 0.006241$. Again, setting $L = l + \varepsilon$ with $l = (B/X)^{1/2}N$, we have to verify (5.41), which is equivalent to (5.44). This inequality is true for $N \geq 4$ since $l = 0.3533N$. With (5.41) satisfied we have that

$$C_2 \geq (X^{1/2} - B^{1/2})^2 \geq 0.0209. \quad (5.48)$$

This proves Theorem 3 for $N \geq 4$. When $N = 1$ there is nothing to prove, while for $N = 2$ we require

$$p_1^2 + p_2^2 \geq 0.0209q^{-2/3}|x_1 - x_2|^{-2}. \quad (5.49)$$

Since $p_1^2 \geq \frac{1}{4}|x_1 - x_2|^{-2}$ for all x_2 , inequality (5.49) is satisfied. For $N = 3$ it suffices to have

$$p_1^2 \geq 0.0209q^{-2/3}\{|x_1 - x_2|^{-2} + |x_1 - x_3|^{-2}\}, \quad (5.50)$$

and this is clearly true by the inequality just mentioned. $\square$

Remarks. In the above proof, the inequality for p^2 was proved in a fashion analogous to that for $|p|$ by substituting Theorem 12 for Theorem 7. However, another proof for p^2 can be given by using the IMS localization [see (3.3)] instead of Theorem 12.

VI. Refined Localization of the Kinetic Energy

Proof of Theorem 10 (Sect. III). Starting from the Corollary of Theorem 9, we see from (3.10) that our task is to find an upper bound to $\text{Tr} \gamma L$ with $L = L^0 + L_1^*$ and with

$$L(x, y) = \pi^{-2}|x - y|^4[1 - \chi_0(x)\chi_0(y) - \chi_1(x)\chi_1(y)] \quad (6.1)$$

and

$$L_1^*(x, y) = \begin{cases} L(x, y)B_1(x)B_1(y) & \text{if } |x - y| \leq \sigma \\ 0 & \text{if } |x - y| > \sigma. \end{cases} \quad (6.2)$$

Recall that B_1 is a ball of radius D_1 centered at the origin. By simple scaling we can, and shall take $D_1 = 1$; we shall also write $B_1 = B$. We have $\chi_1(x) = 0$ unless $|x| \leq (1 - \sigma)$, i.e. unless $x \in B^{(\sigma)}$.

We first bound $\text{Tr} \gamma L^0$. Notice that when $|x| < |y|$, $L^0(x, y) = 0$ unless $|x| \leq (1 - \sigma)$. Using the symmetry of L_0 we can write

$$\text{Tr} \gamma L^0 = 2\text{Re} \iiint_{|x| < |y|} \gamma^{1/2}(x, z)\gamma^{1/2}(z, y)L^0(x, y)B^{(\sigma)}(x)dx dy dz, \quad (6.3)$$

where $\gamma^{1/2}$ is the operator square root of γ . We do the y integration first and then apply Minkowski's inequality to the x integration. For any $\varepsilon > 0$,

$$\begin{aligned} \text{Tr} \gamma L^0 &\leq \varepsilon \iint |\gamma^{1/2}(x, z)|^2 B^{(\sigma)}(x) dx dz \\ &+ \varepsilon^{-1} \iint \left| \int_{|y| \geq |x|} \gamma^{1/2}(z, y) L^0(x, y) dy \right|^2 B^{(\sigma)}(x) dx dz. \end{aligned} \quad (6.4)$$

The first integral is just

$$\int \gamma(x, x) B^{(\sigma)}(x) dx. \quad (6.5)$$

In the second integral we do the z integration before the x integration and obtain

$$\iint \gamma(y, y') \left(\int_A L^0(x, y) L^0(x, y') dx \right) dy dy', \quad (6.6)$$

where A is the region $|x| \leq \min((1 - \sigma), |y|, |y'|)$. The factor in parentheses in (6.6) is the kernel of a positive definite operator, so we can bound (6.6) by

$$\|\gamma\| \int_A L^0(x, y)^2 dx dy, \quad (6.7)$$

where A is the region $|x| \leq (1 - \sigma)$ and $|y| \geq |x|$. In view of the fact that $L^0(x, y)$ is symmetric and $L^0(x, y) = 0$ unless at least one of $|x|$ or $|y|$ is less than $(1 - \sigma)$, and given that $\|\gamma\| = q$ by assumption, (6.7) is just $\frac{1}{2} q \text{Tr}(L^0)^2$. Thus,

$$\text{Tr} \gamma L^0 \leq \varepsilon \int \gamma(x, x) B^{(\sigma)}(x) dx + q \varepsilon^{-1} \Omega_1 \quad (6.8)$$

with $\Omega_1 = \frac{1}{2} \text{Tr}(L^0)^2$. The verification of the two integrals for Ω_1 in (3.15), (3.16) is evident if one recognizes that $\chi_0(x) = 1$ and $\chi_1(x) = 0$ for $|x| \geq (1 - \sigma)$.

Now we turn to $\text{Tr} \gamma L_1^*$. Since γ is a positive operator, its kernel satisfies $|\gamma(x, y)|^2 \leq \gamma(x, x) \gamma(y, y)$. Hence, since $L_1^*(x, y) > 0$ and $h_1(x) > 0$,

$$\begin{aligned} \text{Tr} \gamma L_1^* &= \iint \gamma(x, y) L_1^*(x, y) dx dy \\ &\leq \iint [\gamma(x, x) h_1(y)/h_1(x)]^{1/2} [\gamma(y, y) h_1(x)/h_1(y)]^{1/2} L_1^*(x, y) dx dy \\ &\leq \iint [\gamma(x, x) h_1(y)/h_1(x)] L_1^*(x, y) dx dy \\ &= \int \gamma(x, x) \theta_1(x) dx. \end{aligned} \quad (6.9)$$

The second inequality in (6.9) is the Schwarz inequality, together with the symmetry in x and y . The idea of using the Schwarz inequality in this fashion goes back to Hardy and Littlewood; see [18] for another application.

When inequalities (6.8) and (6.9) are inserted into (3.10), the Corollary of Theorem 9, the result is Theorem 10. $\square$

VII. Estimates of Negative Eigenvalues

Proof of Theorem 11 (Sect. III). It obviously suffices to consider the case $q = 1$. Let the kernel of γ be

$$\gamma(x, y) = \sum_{\alpha} \tau_{\alpha} f_{\alpha}(x) \overline{f_{\alpha}(y)} \quad (7.1)$$

with $0 \leq \tau_{\alpha} \leq 1$ and $\sum \tau_{\alpha} < \infty$ and with the f_{α} being orthonormal. Let $g_{\alpha}(x) \equiv \chi(x) f_{\alpha}(x)$. We want to prove that, with $V(x) = 2/(\pi|x|) + C/R$,

$$E \equiv \sum_{\alpha} (g_{\alpha}, (|p| - V) g_{\alpha}) \geq -4.4827(3/4\pi R^3) C^4 R^{-1} \|\chi\|_2^2. \quad (7.2)$$

By scaling it clearly suffices to prove the theorem for $R=1$, which we assume henceforth.

It is convenient to use Fourier transforms. Let

$$\varrho(p, q) = \iint \bar{\chi}(x) \chi(y) \gamma(x, y) \exp(ip \cdot x - iq \cdot y) dx dy. \quad (7.3)$$

Since $\bar{\chi}\gamma\chi$ is positive semidefinite, so is ϱ , and hence

$$|\varrho(p, q)| \leq \varrho(p, p)^{1/2} \varrho(q, q)^{1/2} \equiv \mu(p) \mu(q) \quad (7.4)$$

with $\mu(p) = \varrho(p, p)^{1/2}$. From (7.3) and the fact that $0 \leq \gamma \leq 1$ as an operator,

$$\mu(p)^2 = (n_p, \gamma n_p) \leq (n_p, n_p) = \int |\chi(x)|^2 dx \equiv M^2, \quad (7.5)$$

where $n_p(x) = \chi(x) \exp(-ip \cdot x)$ and $M = \|\chi\|_2$. Using the Fourier transform of $|x|^{-1}$, namely

$$4\pi |p|^{-2} = \int |x|^{-1} \exp(ip \cdot x) dx, \quad (7.6)$$

E can be written as

$$E = (2\pi)^{-3} \left\{ \int \varrho(p, p) (|p| - C) dp - \pi^{-3} \iint \varrho(p, q) |p - q|^{-2} dp dq \right\}. \quad (7.7)$$

Using (7.5) we have that

$$E \geq (2\pi)^{-3} \inf \{ \tilde{E}(\mu) | 0 \leq \mu(p) \leq M \text{ for all } p \}, \quad (7.8)$$

where $\tilde{E}(\mu)$ is defined by

$$\tilde{E}(\mu) = \int \mu(p)^2 (|p| - C) dp - \pi^{-3} \iint \mu(p) \mu(q) |p - q|^{-2} dp dq. \quad (7.9)$$

To bound the second integral in (7.9), let

$$h(p) = \begin{cases} A^{-2} & \text{if } |p| \leq A \\ |p|^{-2} & \text{if } |p| > A, \end{cases} \quad (7.10)$$

where A is some constant to be determined later. Employing the same strategy as in (6.9) we have

$$\begin{aligned} & \iint \mu(p) \mu(q) |p - q|^{-2} dp dq \\ &= \int \mu(p) (h(q)/h(p))^{1/2} \mu(q) (h(p)/h(q))^{1/2} |p - q|^{-2} dp dq \leq \int \mu(p)^2 t(p) dp, \end{aligned} \quad (7.11)$$

with

$$\begin{aligned} t(p) &= h(p)^{-1} \int |p - q|^{-2} h(q) dq \\ &= h(p)^{-1} \left\{ \int |p - q|^{-2} q^{-2} dq - s(p) \right\} = h(p)^{-1} \left\{ \pi^3 |p|^{-1} - s(p) \right\}, \end{aligned} \quad (7.12)$$

and with

$$s(p) = \int_{|q| < A} |p - q|^{-2} (q^{-2} - A^{-2}) dq. \quad (7.13)$$

To calculate $s(p)$ we use bipolar coordinates, i.e. for any functions f and g

$$\int f(|p - q|) g(|q|) d^3 q = (2\pi/|p|) \int_0^\infty \beta f(\beta) \left\{ \int_{||p|-\beta|}^{|p|+\beta} \alpha g(\alpha) d\alpha \right\} d\beta. \quad (7.14)$$

Thus,

$$\begin{aligned} s(p) &= (2\pi/|p|) \int_0^A (\beta^{-1} - \beta A^{-2}) \left\{ \int_{||p|-\beta|}^{|p|+\beta} \alpha^{-1} d\alpha \right\} d\beta \\ &= (2\pi/|p|) \int_0^{1/\xi} (u^{-1} - u\xi^2) \ln \left(\frac{1+u}{|1-u|} \right) du \end{aligned} \quad (7.15)$$

with $\xi = |p|/A$.

We claim that

$$s(p) \geq \tilde{s}(p) \equiv \begin{cases} (8\pi/3)A & \text{for } |p| \geq A \\ 4\pi|p|^{-1} \left[\frac{10}{9} + \frac{\pi^2}{8} - 2\xi + \frac{1}{6}\xi^2 + \frac{5}{36}\xi^3 \right] & \text{for } |p| \leq A. \end{cases} \quad (7.16)$$

We shall prove (7.16) later. For now, let us insert (7.16) into (7.12), and then into (7.11) and (7.9),

$$\begin{aligned} \tilde{E}(\mu) &\geq \int_{|p| > A} \mu(p)^2 [8A(3\pi^2)^{-1} - C] dp + \int_{|p| < A} \mu(p)^2 [|p| - A^2|p|^{-1} \\ &\quad + \pi^{-3} A^2 \tilde{s}(p) - C] dp. \end{aligned} \quad (7.17)$$

We choose

$$A = 3\pi^2 C/8 \quad (7.18)$$

so that the first integral in (7.17) vanishes. Then, using (7.18) and performing the angular integration,

$$\begin{aligned} \tilde{E}(\mu) &\geq 4\pi A^4 \int_0^1 \mu(Aw)^2 \left\{ w + 2w^2/3\pi^2 + 5w^3/9\pi^2 \right. \\ &\quad \left. - \left[\frac{1}{2} - 40/(9\pi^2) \right] w^{-1} - 32/(3\pi^2) \right\} w^2 dw. \end{aligned} \quad (7.19)$$

As is easily seen, the factor $\{ \}$ in (7.19) has its maximum at $w=1$ and it is negative there. Therefore the infimum of the right side of (7.19) over the set $\mu(Aw) \leq M$ occurs for $\mu(Aw) = M$ for all $0 \leq w \leq 1$. The right side of (7.19) with $\mu = M$ is

$$-(598/135\pi) A^4 M^2. \quad (7.20)$$

Returning to (7.8) and using (7.18) and (7.20) (with 598 replaced by 600) we have that

$$E \geq -\frac{5}{9} \left(\frac{3\pi}{8} \right)^4 C^4 M^2. \quad (7.21)$$

Since $M = \|\chi\|_2$, (7.21) is the same as (7.2).

To complete the proof we must bound (7.15) by (7.16). When $u \leq 1/\xi$, the factor $u^{-1} - u\xi^2 \geq 0$. When $\xi \geq 1$ (i.e. $|p| \geq A$), $u \leq 1$ and we have the bound

$$\ln[(1+u)/(1-u)] \geq 2u. \quad (7.22)$$

Inserting (7.22) into (7.15) yields the first part of (7.16).

If $|p| \leq A$, then $\xi < 1$. The integral in (7.15) from 0 to 1 can be done explicitly,

$$\int_0^1 (u^{-1} - u\xi^2) \ln[(1+u)/(1-u)] du = \pi^2/4 - \xi^2. \quad (7.23)$$

To bound the integral from 1 to $1/\xi$, use the fact that for $u > 1$,

$$\ln[(1+u)/(u-1)] \geq 2u^{-1} + \frac{2}{3}u^{-3}. \quad (7.24)$$

Then

$$\begin{aligned} & \int_1^{1/\xi} (u^{-1} - u\xi^2) \ln[(1+u)/(u-1)] du \\ & \geq \int_1^{1/\xi} (u^{-1} - u\xi^2) \left(2u^{-1} + \frac{2}{3}u^{-3} \right) du = 20/9 - 4\xi + 4\xi^2/3 + 4\xi^3/9. \end{aligned} \quad (7.25)$$

When (7.25) is combined with (7.23) (and the $4\xi^3/9$ term is replaced by the smaller quantity $5\xi^3/18$) the result is the second part of (7.16). $\square$

VIII. Some Numerical Calculations

Our goal here is to derive the bounds (3.33) for Ω and (3.34) for $\theta(r)$.

(A) *Evaluation of Ω .* Ω is defined as the sum of the two integrals in (3.15), (3.16). Recall that $\sigma = 0.3$ and $\chi_1(x) = \chi(|x|)$ is given in (3.22) while $\chi_0(x)^2 = 1 - \chi_1(x)^2$. We already set $D_1 = 1$.

To evaluate $I^{(1)}$ we use the spherical symmetry of χ and first do the angular integration on x and y . This integral is

$$\begin{aligned} \int |x-y|^{-8} d\omega_y &= 2\pi \int_0^\pi (x^2 + y^2 - 2xy \cos \theta)^{-4} \sin \theta d\theta \\ &= (\pi/3) (|x||y|)^{-1} \{ (|x| - |y|)^{-6} - (|x| + |y|)^{-6} \}. \end{aligned} \quad (8.1)$$

Thus,

$$\begin{aligned} I^{(1)} &= 4(3\pi^2)^{-1} \int_0^{1-2\sigma} ds \int_{s+\sigma}^{1-\sigma} dt [(t-s)^{-6} \\ &\quad - (t+s)^{-6}] [1 - (1 - \chi(s)^2)^{1/2} (1 - \chi(t)^2)^{1/2} - \chi(s)\chi(t)]^2. \end{aligned} \quad (8.2)$$

(Note that we integrate over $t > s + \sigma$ and $s, t < 1 - \sigma$, and then multiply by 2. Since $s < t - \sigma$ and $t < 1 - \sigma$, we have that $s < 1 - 2\sigma$.) This integral is not elementary, but because it is an integral of a continuous, bounded function over a bounded domain in $\mathbb{R}^2$ it can be confidently evaluated on a computer. The result is (3.33).

To evaluate $I^{(2)}$, the angular integration over y is done first as before, with the result (8.1). Then $I^{(2)}$ is the sum of three integrals according as $|x| < 1 - 3\sigma$, $1 - 3\sigma \leq |x| < 1 - 2\sigma$, $1 - 2\sigma \leq |x| \leq 1 - \sigma$. Thus,

$$\begin{aligned} I^{(2)} &= (4/3\pi^2) \int_0^{1-3\sigma} ds \int_{1-\sigma}^\infty dt [(t-s)^{-6} - (t+s)^{-6}] \\ &\quad + (16/3\pi^2) \int_{1-3\sigma}^{1-2\sigma} ds \int_{1-\sigma}^\infty dt [(t-s)^{-6} - (t+s)^{-6}] \sin^4 \left[\frac{\pi}{8\sigma} (1 - \sigma - s) \right] \\ &\quad + (16/3\pi^2) \int_{1-2\sigma}^{1-\sigma} ds \int_{s+\sigma}^\infty dt [(t-s)^{-6} - (t+s)^{-6}] \sin^4 \left[\frac{\pi}{8\sigma} (1 - \sigma - s) \right]. \end{aligned} \quad (8.3)$$

In each case the t integration can easily be done analytically. This transforms (8.3) into three integrals over the bounded intervals $0 \leq s < 1 - 3\sigma$, $1 - 3\sigma \leq s < 1 - 2\sigma$ and $1 - 2\sigma \leq s \leq 1 - \sigma$. The integrands are again bounded and continuous so numerical integration can be used. The result is (3.33).

(B) *Bound on $\theta(r)$, Eq. (3.34).* The function $\theta \equiv \theta_1$ is defined in (3.13) with h defined in (3.24). Again we take $D_1 = 1$. The kernel L_1^* is given in (3.12) with $\chi_1 \equiv \chi$ given in (3.22) and $\chi_0^2 \equiv 1 - \chi^2$.

We want to compute

$$I(r) = \int L_1^*(x, y) h(|y|) dy \quad (8.4)$$

with $r = |x|$. Since the angular integral of $|x - y|^{-4}$ is less than $\pi(rs)^{-1}(r - s)^{-2}$, with $s = |y|$, we have that

$$I(r) \leq (1/\pi r) \int_0^1 (r - s)^{-2} h(s) m(r, s) ds, \quad (8.5)$$

where $m(r, s) = m(s, r)$ and, for $r \leq s$, $m(r, s)$ is given by

$$m(r, s) = \begin{cases} 1 - \cos[\pi(s - \tau)/4\sigma] & \text{for } 0 \leq r \leq \tau \leq s \leq r + \sigma \\ 1 - \cos[\pi(s - r)/4\sigma] & \text{for } \tau \leq r \leq s \leq \min(\tau + 2\sigma, r + \sigma) \\ 1 - \cos[\pi(2\sigma + \tau - r)/4\sigma] & \text{for } s - \sigma \leq r \leq \tau + 2\sigma \leq s \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (8.6)$$

In (8.6), $\tau = 1 - 3\sigma$.

The arguments of the cosines in (8.6) are all at most $\pi/4$ and one can use the inequality $\cos b \geq 1 - b^2/2$ for $|b| \leq \pi/4$. If we use this inequality in (8.6) and then insert the result in (8.5), the integral (8.5) is seen to be elementary but tedious [recall (3.24)]. Finally, $\theta(r) = I(r)/h(r)$.

Let us verify (3.34) when $r \geq 1 - \sigma$. Then $rh(r) = 1$ and thus

$$\theta(r) = (\pi/32\sigma^2) \int_{r-\sigma}^{1-\sigma} sh(s)(r-s)^{-2}(1-\sigma-s)^2 ds. \quad (8.7)$$

The second line of (3.24) is appropriate for this region. In the region $r - \sigma \leq s \leq 1 - \sigma$ the function $(r-s)^{-2}(1-\sigma-s)^2$ is monotone decreasing in s and so has its maximum at $s = r - \sigma$. Thus,

$$\theta(r) \leq (\pi/32\sigma^2) \sigma^{-2} (1-r)^2 \int_{r-\sigma}^{1-\sigma} \{2 - \sigma(s - 1 + 2\sigma)\} ds, \quad (8.8)$$

and this agrees with (3.34) for $r \geq 1 - \sigma$.

The verification of the $r \leq 1 - \sigma$ case of (3.34) is elementary and we omit the details.

IX. The Occurrence of Collapse for Large α

In the previous sections it was shown that the Hamiltonian H_{NK} (1.4) under consideration is stable if α is small enough. There are two parameters in the problem, $z\alpha$ and α . For stability of one electron and one nucleus it is necessary and

sufficient that $z\alpha \leq 2/\pi$, but, assuming this condition, there is stability in the many-body case if $\alpha < \alpha_0/q$ with $\alpha_0 > 1/47$. In this section we shall prove that this stability bound is not just an artifact of our proof but that instability definitely occurs if α is too large. Theorems 3 and 4 will be proved here.

Proof of Theorem 3. The method of proof here is the same as the method employed in [23] to prove the instability of one-electron molecules in a magnetic field. Let $\phi \in L^2(\mathbb{R}^3)$ be real with $\|\phi\|_2 = 1$ and let $\tau = (\phi, |p|\phi)$ which is assumed to be finite. Then

$$E = (\phi, H_{NK}\phi) = \tau - z\alpha \int \phi^2(x) \sum_{j=1}^K |x - R_j|^{-1} dx + z^2\alpha \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1}. \quad (9.1)$$

With ϕ fixed let us try to position the R_i so as to minimize the right side of (9.1). This minimum (call it e) is less than any average of E over positions of the R_i . In particular, we use $\psi = \prod_{j=1}^K \phi(R_j)^2$ as a probability density for such an average. Then

$$Av(E) = \tau - \sigma [z\alpha K - z^2\alpha K(K-1)/2] = \tau + \frac{1}{2}\sigma \{z^2\alpha [K - \frac{1}{2} - z^{-1}]^2 - \frac{1}{4}z^2\alpha - \alpha - z\alpha\}, \quad (9.2)$$

where

$$\sigma = \int \phi(x)^2 \phi(y)^2 |x - y|^{-1} dx dy. \quad (9.3)$$

Now K can be chosen so that $|K - \frac{1}{2} - z^{-1}| \leq \frac{1}{2}$. Using this K , we have

$$e \leq Av(E) \leq \tau - \frac{1}{2}\sigma\alpha. \quad (9.4)$$

If we set $\alpha_1 = 2\tau/\sigma$, then when $\alpha > \alpha_1$, $e < 0$, and we can drive e to $-\infty$ simply by dilation, i.e. $\phi(x) \rightarrow \lambda^{3/2}\phi(\lambda x)$ and $R_j \rightarrow \lambda R_j/\lambda$ with $\lambda \rightarrow \infty$.

To obtain a numerical value for α_1 , choose $\phi(x) = \pi^{-1/2} \exp(-r)$ with $r = |x|$. The Fourier transforms of ϕ and ϕ^2 are

$$\hat{\phi}(p) = 8\pi^{1/2}(1 + p^2)^{-2}, \quad \widehat{\phi^2}(p) = 16(4 + p^2)^{-2}. \quad (9.5)$$

Then

$$\tau = (2\pi)^{-3} \int \hat{\phi}(p)^2 |p| dp = 8/3\pi, \quad \sigma = (2\pi)^{-3} \int \widehat{\phi^2}(p) (4\pi/|p|^2) dp = 5/8, \quad (9.6)$$

and $2\tau/\sigma = 128/15\pi$. $\square$

Proof of Theorem 4. The method of proof here is similar to that used in [20] to prove that the energy of N nonrelativistic bosons interacting with fixed nuclei via Coulomb forces diverges as $-N^{5/3}$. Again, let $\phi \in L^2(\mathbb{R}^3)$ be real with $\|\phi\|_2 = 1$ and $\tau = (\phi, |p|\phi)$. Since there are q spin states, we can put $N = q$ electrons into the state ϕ . The energy is then

$$E = q\tau - z\alpha q \int \phi^2(x) \sum_{j=1}^K |x - R_j|^{-1} dx + z^2\alpha \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1} + \frac{1}{2} q(q-1)\sigma \quad (9.7)$$

with σ given in (9.3). Let us first prove the theorem under the condition $q/z \geq 1$; at the end of the proof we shall show how to handle the case $q/z < 1$.

To construct ϕ we first define $g \in L^2(\mathbb{R}^3)$ by

$$g(x, y, z) = f(x) f(y) f(z), \quad (9.8)$$

where $f \in L^2(\mathbb{R}^1)$ is given by $f(x) = \sqrt{3/2}(1 - |x|)$ for $|x| \leq 1$ and $f(x) = 0$ for $|x| \geq 1$. This f has $\|f\|_2 = 1$, and thus $\|g\|_2 = 1$. Let $h \in L^2(\mathbb{R}^3)$ be some other function with compact support and with $(h, |p|h) < \infty$ and $\|h\|_2 = 1$. Define the integers n and K and the positive number λ by

$$\begin{aligned} n &= [(q/z)^{1/3}] \geq 1, \quad K = n^3, \\ \lambda &= n^3 z/q = Kz/q, \end{aligned} \quad (9.9)$$

where $[b]$ means integral part of b . Clearly, $1 \geq \lambda \geq 1/8$. Finally, we construct a sequence of functions $\phi^{(s)}(x)$, $x \in \mathbb{R}^3$, by

$$\phi^{(s)}(x)^2 = \lambda g(x) + (1 - \lambda) s^{-3} h(x/s + (0, 0, s^2))^2. \quad (9.10)$$

Now choose some fixed locations $R_1, \dots, R_K$ of K nuclei. Because of the scaling of h by s^{-1} and translation by $(0, 0, s^2)$, we have that E converges to the following E' as $s \rightarrow \infty$:

$$E' = q\lambda\tau - z\alpha\lambda q \int g^2(x) \sum_{j=1}^K |x - R_j|^{-1} dx + z^2\alpha \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1} + \frac{1}{2} \lambda^2 q(q-1)\sigma\alpha, \quad (9.11)$$

where τ now means $(g, |p|g)$ and σ is given in (9.3) with g in place of ϕ .

We claim that it is possible to choose the locations $R_1, \dots, R_K$ so that

$$\sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1} - K \int g^2(x) \sum_{j=1}^K |x - R_j|^{-1} dx + \frac{1}{2} K^2 \sigma \leq -K^{4/3}/6. \quad (9.12)$$

If (9.12) holds then, recalling (9.9),

$$E' \leq q\lambda\tau - \frac{1}{6} z^2 \alpha (\lambda q/z)^{4/3}. \quad (9.13)$$

Recalling that $\lambda > 1/8$ we have that $E' < 0$ whenever

$$\beta^2 \alpha q \geq 8(6\tau)^3 (\pi/2)^2. \quad (9.14)$$

We also have that $\tau = (g, |p|g) \leq (g, p^2 g)^{1/2} = 3$ (by the Schwarz inequality). Thus, collapse occurs if $\alpha \geq \alpha_2 q^{-1} \beta^{-2}$ with $\alpha_2 = (\pi/2)^2 8(18)^3 = 115, 120$, provided $q/z \geq 1$.

If, on the other hand, $q/z < 1$ and if $\alpha > \alpha_2 q^{-1} \beta^{-2} = \alpha_2 q^{-1} z^{-2} \alpha^{-2} (2/\pi)^2$, we have that $(\pi/2)^2 (z\alpha)^3 > \alpha_2 z/q > \alpha_2$. Since $\alpha_2 \gg (2/\pi)^5$, we are in the situation that $z\alpha > 2/\pi$, which certainly entails collapse. Therefore, the theorem is proved for all ratios q/z with the α , given above.

There remains to prove (9.12). Choose $n-1$ numbers $\beta_1, \dots, \beta_{n-1}$ satisfying $-1 \equiv \beta_0 < \beta_1 < \dots < \beta_{n-1} < \beta_n \equiv 1$ such that

$$\int_{\beta_j}^{\beta_{j+1}} f(x)^2 dx = 1/n \quad \text{for all } j.$$

Let L_j be the interval $[\beta_{j-1}, \beta_j]$ in $\mathbb{R}^1$ and, with m denoting a triplet (i, j, k) , let $\Gamma(m) \subset \mathbb{R}^3$ be the rectangular parallelepiped $L_i \times L_j \times L_k$. Then, for each m ,

$$\int_{\Gamma(m)} g^2(x) dx = 1/n^3 = 1/K. \quad (9.15)$$

There are n^3 of these parallelepipeds. To prove (9.12) we shall place one of the R_i 's in each $\Gamma(m)$ and average its location with respect to the density $g^2(x)$ restricted to $\Gamma(m)$. If the average satisfies (9.12) then there is surely some choice of the R_i 's that satisfies (9.12). Apart from a self energy contribution from each parallelepiped, the average of the left side of (9.12) is zero. Thus the average of the left side is given by the self energy terms

$$W = -\frac{1}{2} n^6 \sum_m \int_{\Gamma(m)} \int_{\Gamma(m)} g(x)^2 g(y)^2 |x-y|^{-1} dx dy. \quad (9.16)$$

Each integral is the self energy of a charge density g^2 in $\Gamma(m)$. However $\Gamma(m)$ lies inside a ball $B(m)$ of radius $r(m) = (s^2 + t^2 + u^2)^{1/2}$, where $2s$, $2t$, and $2u$ are the lengths of $\Gamma(m)$, namely $(\beta_i - \beta_{i-1})$, $(\beta_j - \beta_{j-1})$, $(\beta_k - \beta_{k-1})$. The self energy is greater than the *minimum* self energy of a charge $1/K$ distributed in $B(m)$; the minimum occurs for a uniform charge distribution on the boundary of $B(m)$ and is $r(m)^{-1}/K^2$. Thus,

$$W \leq -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (s^2 + t^2 + u^2)^{-1/2}. \quad (9.17)$$

Now $(s^2 + t^2 + u^2)^{-1/2} > (s + t + u)^{-1}$. Substituting this latter expression in (9.17) and then using the convexity of the function $(s, t, u) \rightarrow (s + t + u)^{-1}$ and recalling that $K = n^3$, we have that

$$W \geq -\frac{1}{2} K(a + b + c)^{-1}, \quad (9.18)$$

where a , b , and c are the averages of s , t , and u . But $a = b = c = 1/n$, and thus (9.12) is proved. $\square$

Acknowledgements. The authors thank Michael Loss for helpful discussions and comments and they thank Stefan Knabe for performing numerical calculations.

References

1. Baxter, J.R.: Inequalities for potentials of particle systems. Ill. J. Math. **24**, 645–652 (1980)
2. Chandrasekhar, S.: Phil. Mag. **11**, 592 (1931); Astro. J. **74**, 81 (1931); Monthly Notices Roy. Astron. Soc. **91**, 456 (1931); Rev. Mod. Phys. **56**, 137 (1984)
3. Conlon, J.G.: The ground state energy of a classical gas. Commun. Math. Phys. **94**, 439–458 (1984)
4. Conlon, J.G., Lieb, E.H., Yau, H.-T.: The $N^{7/5}$ law for charged bosons. Commun. Math. Phys. **116**, 417–448 (1988)
5. Cycon, H.L., Froese, R.G., Kirsch, W., Simon, B.: Schrödinger operators. Berlin, Heidelberg, New York: Springer 1987
6. Daubechies, I.: An uncertainty principle for fermions with generalized kinetic energy. Commun. Math. Phys. **90**, 511–520 (1983)
7. Daubechies, I.: One electron molecules with relativistic kinetic energy: properties of the discrete spectrum. Commun. Math. Phys. **94**, 523–535 (1984)
8. Daubechies, I., Lieb, E.H.: One-electron relativistic molecules with Coulomb interaction. Commun. Math. Phys. **90**, 497–510 (1983)
9. Dyson, F.J.: Ground state energy of a finite system of charged particles. J. Math. Phys. **8**, 1538–1545 (1967)

10. Dyson, F.J., Lenard, A.: Stability of matter I and II. *J. Math. Phys.* **8**, 423–434 (1967); *ibid* **9**, 698–711 (1968). See also Lenard's Battelle lecture. In: *Lecture Notes in Physics*, vol. 23. Berlin, Heidelberg, New York: Springer 1973
11. Erdelyi, A., Magnus, W., Oberhettinger, F., Tricomi, F.G.: *Tables of integral transforms*, Vol. 1. New York, Toronto, London: McGraw-Hill 1954, p. 75, 2.4 (35)
12. Federbush, P.: A new approach to the stability of matter problem. II. *J. Math. Phys.* **16**, 706–709 (1975)
13. Fefferman, C.: The N -body problem in quantum mechanics. *Commun. Pure Appl. Math. Suppl.* **39**, S67–S109 (1986)
14. Fefferman, C., de la Llave, R.: Relativistic stability of matter. I. *Rev. Math. Iberoamericana* **2**, 119–215 (1986)
15. Fröhlich, J., Lieb, E.H., Loss, M.: Stability of Coulomb systems with magnetic fields. I. The one-electron atom. *Commun. Math. Phys.* **104**, 251–270 (1986)
16. Herbst, I.: Spectral theory of the operator $(p^2 + m^2)^{1/2} - ze^2/r$. *Commun. Math. Phys.* **53**, 285–294 (1977); Errata *ibid* **55**, 316 (1977)
17. Kato, T.: *Perturbation theory for linear operators*. Berlin, Heidelberg, New York: Springer 1966. See remark 5.12, p. 307
18. Kovalenko, V., Perelmuter, M., Semenov, Ya.: Schrödinger operators with $L_w^{1/2}(\mathbb{R}^1)$ potentials. *J. Math. Phys.* **22**, 1033–1044 (1981)
19. Lieb, E.H.: Stability of matter. *Rev. Mod. Phys.* **48**, 553–569 (1976)
20. Lieb, E.H.: The $N^{5/3}$ law for bosons. *Phys. Lett.* **70A**, 71–73 (1979)
21. Lieb, E.H.: Density functionals for Coulomb systems. *Int. J. Quant. Chem.* **24**, 243–277 (1983)
22. Lieb, E.H.: On characteristic exponents in turbulence. *Commun. Math. Phys.* **92**, 473–480 (1984)
23. Lieb, E.H., Loss, M.: Stability of Coulomb systems with magnetic fields. II. The many electron atom and the one electron molecule. *Commun. Math. Phys.* **104**, 271–282 (1986)
24. Lieb, E., Simon, B.: Thomas Fermi theory of atoms, molecules and solids. *Adv. Math.* **23**, 22–116 (1977)
25. Lieb, E.H., Thirring, W.: Bound for the kinetic energy of fermions which proves the stability of matter. *Phys. Rev. Lett.* **35**, 687–689 (1975). Errata, *ibid* **35**, 1116 (1975); see also their article: Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In: *Studies in Mathematical Physics, Essays in honor of Valentine Bargmann*. Lieb, E.H., Simon, B., Wightman, A.S. (eds.). Princeton, NJ: Princeton University Press 1976
26. Lieb, E.H., Thirring, W.: Gravitational collapse in quantum mechanics with relativistic kinetic energy. *Ann. Phys. (NY)* **155**, 494–512 (1984)
27. Lieb, E.H., Yau, H.-T.: The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics. *Commun. Math. Phys.* **112**, 147–174 (1987). See also Lieb, E.H. and Yau, H.-T.: A rigorous examination of the Chandrasekhar theory of stellar collapse. *Astro. J.* **323**, 140–144 (1987)
28. Loss, M., Yau, H.-T.: Stability of Coulomb systems with magnetic fields. III. Zero energy bound states of the Pauli operator. *Commun. Math. Phys.* **104**, 283–290 (1986)
29. Weder, R.: Spectral analysis of pseudodifferential operators. *J. Funct. Anal.* **20**, 319–337 (1975)

Communicated by A. Jaffe

Received May 12, 1988

Stability of Relativistic Matter via Thomas–Fermi Theory

By Elliott H. Lieb¹, Michael Loss² and Heinz Siedentop³

⁽¹⁾ Departments of Mathematics and Physics, Princeton University,
P.O. Box 708, Princeton, NJ 08544-0708, USA

⁽²⁾ Georgia Institute of Technology, Atlanta, GA 30332-0160, USA

⁽³⁾ Matematisk institutt, Universitetet i Oslo, Postboks 1053, N-0316 Oslo, Norway

Abstract A Thomas-Fermi-Weizsäcker type theory is constructed, by means of which we are able to give a relatively simple proof of the stability of relativistic matter. Our procedure has the advantage over previous ones in that the lower bound on the critical value of the fine structure constant, α , is raised from 0.016 to 0.77 (the critical value is known to be less than 2.72). When $\alpha = 1/137$, the largest nuclear charge is 59 (compared to the known optimum value 87). Apart from this, our method is simple, for it parallels the original Lieb-Thirring proof of stability of nonrelativistic matter, and it adds another perspective on the subject.

This article is dedicated to our colleagues, teachers, and coauthors Klaus Hepp and Walter Hunziker on the occasion of their sexagesimal birthdays. Their enthusiasm for quantum mechanics as an unending source of interesting physics and mathematics has influenced many.

¹ Work partially supported by U.S. National Science Foundation grant PHY95-13072.

² Work partially supported by U.S. National Science Foundation grant DMS95-00840.

³ Work partially supported by European Union, grant ERBFMRXCT960001.

©1996 by the authors. Reproduction of this article, in its entirety, by any means is permitted for non-commercial purposes provided that full reference to the original source of publication is made.

This paper appeared in *Helvetica Physica Acta* vol. 69, no. 5/6, 974-984 (1996). Three typographical errors that appeared in the original paper and in the second edition of this ‘Selecta’ have been corrected here.

1. Introduction

The ‘stability of relativistic matter’ concerns the N -body Hamiltonian (in units of $\hbar c$)

$$H = \sum_{i=1}^N |p_i| + \alpha V_c, \quad (1.1)$$

where V_c is the Coulomb potential of K fixed nuclei with nuclear charge Ze , with locations R_j in $\mathbf{R}^3$, and with N electrons. The Coulomb potential is

$$V_c = -V + R + U, \quad (1.2)$$

where

$$V := Z \sum_{i=1}^N \sum_{j=1}^K |x_i - R_j|^{-1}, \quad (1.3)$$

$$R := \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}, \quad (1.4)$$

$$U := Z^2 \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1}. \quad (1.5)$$

As usual $p = -i\nabla$ and $|p| = \sqrt{-\Delta}$, and the x_j are the electron coordinates. The electrons are assumed to have q spin states each, $q = 2$ being the physical value. This means that the Hilbert space for the N -electron functions is the N -fold antisymmetric tensor product of $L^2(\mathbf{R}^3; \mathbf{C}^q)$. The constant $\alpha = e^2/\hbar c$ is called the fine structure constant.

We can easily include a magnetic field, which means replacing $|p_i|$ by $|p_i + A(x_i)|$. The vector field, A , is the vector potential (in suitable units) of a magnetic field, $B = \text{curl} A$, and can be arbitrary, as far as the present work is concerned. A mass can be included as well, i.e., $|p_i + A(x_i)|$ can be replaced by $\sqrt{|p_i + A(x_i)|^2 + m^2} - m$. The inclusion of a mass or magnetic field, while it changes the energy, does not affect stability. The reason for this and the requisite changes will be pointed out in the final section. It is for simplicity and clarity that we set $m = 0$ and $A = 0$.

‘Stability of matter’ means that the operator, H , is bounded below by a universal constant times $N + K$, independent of the R_j and A . In our case, because everything scales as an inverse length, the lower bound for H is either $-\infty$ or 0. Thus, we have to find the conditions under which H is a positive operator.

Many people worked on various aspects of this problem, including J. Conlon (who gave the first proof [C84]), I. Daubechies, C. Fefferman, I. Herbst, T. Kato, E. Lieb, R. de la Llave, R. Weder, and H-T. Yau. A careful, and still current, review of the history is contained in the introduction to [LY88], to which we refer the reader. For present purposes it suffices to note the current state of affairs concerning the best available constants needed for stability, as derived in [LY88]. We can list these in a sequence of remarks as follows:

1. Stability for any given values, α_* and Z_* , implies stability for all $0 \leq \alpha < \alpha_*$ and $Z < Z_*$. In fact, we can allow the nuclei to have different charges Z_i , $1 \leq i \leq K$, provided $Z_i \leq Z_*$ for all i . This follows from some simple concavity considerations and has nothing to do with the nature of the proof leading to α_* and Z_* .
2. Theorem 2 of [LY88] has the strongest results, but it is limited to the case of zero magnetic field, $A = 0$. The result is that stability occurs if

$$q\alpha \leq 1/47 \quad \text{and} \quad Z\alpha \leq 2/\pi . \quad (1.6)$$

It is not clear to us how to incorporate a magnetic field in the proof of Theorem 2, and we leave this as an open problem.

3. Theorem 1 of [LY88] has weaker results, but a simpler proof. That proof generalizes easily to the $A \neq 0$ case, as pointed out in [LLS95]. The result is complicated to state in full generality, but a representative example is that stability holds if

$$q\alpha \leq 0.032 \quad \text{and} \quad Z\alpha \leq 1/\pi . \quad (1.7)$$

It is possible to let $Z\alpha \rightarrow 2/\pi$ at the expense of $q\alpha \rightarrow 0$.

4. Instability definitely occurs if $Z\alpha > 2/\pi$, or if $Z_i\alpha > 2/\pi$ for any i . It also occurs if

$$\alpha > 128/(15\pi) \approx 2.72 \quad (1.8)$$

for *any* positive value of Z and any value of q . In other words, if $\alpha > 128/(15\pi)$ and if $Z > 0$ then one can produce collapse with only one electron, $N = 1$, by utilizing sufficiently many nuclei, i.e., by choosing K sufficiently large.

5. Instability also definitely occurs if ([LY88], Theorem 4)

$$\alpha > 36q^{-1/3}Z^{-2/3} , \quad (1.9)$$

which implies that bosonic matter (which can always be thought of as fermionic matter with $q = N$) is *always* unstable. (Note: there is a typographical error in Theorem 4 of [LY88].)

2. Main Results

The proof of the stability of *nonrelativistic* matter in [LT75] uses a series of inequalities to relate the ground state energy of the Hamiltonian to the Thomas–Fermi energy of the electron density, $\rho(x)$. The chief point is the kinetic energy inequality for an N -electron state Ψ , namely

$$\langle \Psi | \sum_{i=1}^N |p_i|^2 | \Psi \rangle > \text{const.} \int \rho^{5/3} .$$

The same approach will not work in the relativistic case because the corresponding inequality [D83] is, for dimensional reasons,

$$\langle \Psi | \sum_{i=1}^N |p_i| | \Psi \rangle > \text{const.} \int \rho^{4/3} .$$

While $\int \rho^{5/3}$ can control the Coulomb attraction $-Z\alpha \int \rho(x)/|x|$, unfortunately $\int \rho^{4/3}$ cannot do so. For this reason no attempt seems to have been made to imitate the proof in [LT75] of stability in the relativistic case.

However, the Coulomb singularity *can* be controlled by a Weizsäcker type term, namely $(\sqrt{\rho}, |p| \sqrt{\rho})$. The relativistic kinetic energy can, in turn, be bounded below by a term of this type plus a term of the $\int \rho^{4/3}$ type. This and other essential inequalities will be explained more fully below. With the ‘Coulomb tooth’ now gone, TF theory with $\int \rho^{4/3}$ can deal adequately with the rest of the Coulomb energy (with the aid of the exchange-correlation energy inequality [LO81], whose remainder term also has the form $\int \rho^{4/3}$).

Before going into details, let us state our main results. First, we define Thomas-Fermi-Weizsäcker (TFW) theory as follows: The class of functions (‘densities’) to be considered, denoted by $\mathcal{C}$, consists of those nonnegative functions $\rho : \mathbf{R}^3 \rightarrow \mathbf{R}^+$ such that $\sqrt{\rho}$ and $\sqrt{|p|\rho}$ have finite $L^2(\mathbf{R}^3)$ norms, i.e.,

$$\mathcal{C} = \left\{ \rho : \rho(x) \geq 0 \text{ and } \int_{\mathbf{R}^3} (1 + |p|) |\widehat{\sqrt{\rho}}(p)|^2 dp < \infty \right\} , \quad (2.1)$$

where $\widehat{\sqrt{\rho}}(p) := (2\pi)^{-3/2} \int_{\mathbf{R}^3} \exp[-ip \cdot x] \sqrt{\rho}(x) dx$ denotes the Fourier transform of the function $\sqrt{\rho}(x)$.

Next, we define the functional

$$T(\rho) := \int_{\mathbf{R}^3} |p| |\widehat{\sqrt{\rho}}(p)|^2 dp \equiv (\sqrt{\rho}, |p| \sqrt{\rho}) . \quad (2.2)$$

The TFW functional, with arbitrarily given positive constants β and γ , is then

$$\mathcal{E}(\rho) := \beta T(\rho) + \frac{3}{4} \gamma \int_{\mathbf{R}^3} \rho^{4/3}(x) dx - \alpha \int_{\mathbf{R}^3} V(x) \rho(x) dx + \alpha D(\rho, \rho) + \alpha U \quad (2.3)$$

with

$$D(\rho, \rho) := (1/2) \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \rho(x) \rho(y) |x - y|^{-1} dx dy .$$

The quantity of principal interest is the energy

$$E^{TFW} := \inf \{ \mathcal{E}(\rho) : \rho \in \mathcal{C} \} . \quad (2.4)$$

This quantity depends on the parameters α, β and γ and on the nuclear coordinates, R_j . If, however, we try to minimize E over all choices of the nuclear coordinates then the result is either 0 or $-\infty$, as can be easily seen from the fact that all the terms in $\mathcal{E}$ scale, under dilation, as an inverse length.

THEOREM 1. (Stability of TFW theory). *The TFW energy, E^{TFW} , in (2.4) is nonnegative if*

$$\beta \geq \frac{\pi}{2} Z\alpha, \quad \text{and} \quad \gamma \geq 4.8158 Z^{2/3} \alpha \quad (2.5)$$

On the other hand, if $\beta < (\pi/2)Z\alpha$ then $E = -\infty$ for every choice of the nuclear coordinates.

For the next theorem we have to define the density corresponding to an N -body wave function. If Ψ is an antisymmetric function of N space-spin coordinates, normalized to unity in the usual way, we define

$$\rho_\Psi(x) := N \sum_{1 \leq \sigma_1, \dots, \sigma_N \leq q} \int_{\mathbf{R}^{3(N-1)}} |\Psi(x, \sigma_1; x_2, \sigma_2; \dots; x_N, \sigma_N)|^2 dx_2 \cdots dx_N. \quad (2.6)$$

THEOREM 2. (TFW theory bounds quantum mechanics). *Let Ψ be any normalized antisymmetric function, with ρ_Ψ defined in (2.6). Choose*

$$\beta = \frac{\pi}{2} Z\alpha \quad \text{and} \quad \gamma = \frac{4}{3} \left[1.63q^{-1/3} \left(1 - \frac{\pi}{2} Z\alpha \right) - 1.68 \alpha \right]. \quad (2.7)$$

Assume that γ is positive. Then, with this definition of the TFW functional (2.3),

$$\langle \Psi | H | \Psi \rangle \geq \mathcal{E}(\rho_\Psi). \quad (2.8)$$

A corollary of these two theorems is that *our Hamiltonian, H , in (1.1) is stable if*

$$\left(\frac{\pi}{2}\right)Z + 2.2159 q^{1/3} Z^{2/3} + 1.0307 q^{1/3} \leq 1/\alpha. \quad (2.9)$$

(Cf. (1.9)) In particular, *with $q = 2$ for electrons, relativistic matter is stable if $\alpha < 0.77$ and if Z is not too large. When $\alpha = 1/137$ the allowed Z is 59, which compares favorably with the best possible value $87 \approx 137(2/\pi)$.*

We leave it as a challenge to improve our method so as to achieve the value $137(2/\pi)$ (with a magnetic field present). As noted above, this value has been achieved in [LY88], but without a magnetic field. The most noteworthy point is the large value of the critical fine structure constant: $\alpha_{critical} \geq 0.77$ when $q = 2$.

The bound in (2.9) is, in some respects, similar to Theorem 1 in [LY88], but it is far simpler, clearer and gives the correct q -dependence of α (note that (1.9) gives a similar

bound in the other direction). The chief methodological difference is that Theorem 6 is used in [LY88], which bounds the Coulomb potential below by a one-body potential. Here, we use the exchange-correlation inequality (3.9) instead. We repeat that the results above also hold with a magnetic field.

It is to be emphasized that our stability result is really contained in Theorem 2. Theorem 1 only gives a condition for which $\mathcal{E}(\rho) \geq 0$. A better estimate on the TFW functional will, via Theorem 2, yield a better stability bound.

3. Some Essential Inequalities

There are five known inequalities about Coulomb systems that will be needed in our proof of our main theorems. We begin by recalling them.

KINETIC ENERGY LOCALIZATION, [LY88] pp. 186 and 188.

Denote by Γ_j the Voronoi cell in $\mathbf{R}^3$ that contains R_j , i.e., the set

$$\Gamma_j := \{x \in \mathbf{R}^3 : |x - R_j| \leq |x - R_k| \text{ for all } k\} , \quad (3.1)$$

and let D_j be half the distance of the j -th nucleus to its nearest neighbor. These Γ_j are disjoint, except for their boundaries and, being the intersection of half-spaces they are convex sets. The ball centered at R_j with radius D_j is denoted by B_j . Obviously, $B_j \subset \Gamma_j$.

For any function $f \in L^2(\mathbf{R}^3)$ there is the inequality

$$(f, |p| f) \geq \sum_{j=1}^K \int_{B_j} |f(x)|^2 \left\{ \frac{2}{\pi} |x - R_j|^{-1} - \frac{1}{D_j} Y \left(\frac{|x - R_j|}{D_j} \right) \right\} dx . \quad (3.2)$$

The function Y is given, for $0 < r < 1$, by

$$Y(r) = \frac{2}{\pi(1+r)} + \frac{1+3r^2}{\pi r(1+r^2)} \ln(1+r) - \frac{1-r^2}{\pi r(1+r^2)} \ln(1-r) - \frac{4r}{\pi(1+r^2)} \ln r . \quad (3.3)$$

Numerically it is found that [LY88] (2.27)

$$4\pi \int_0^1 Y(r)^4 r^2 dr < 7.6245 . \quad (3.4)$$

RELATIVISTIC KINETIC ENERGY BOUND FOR FERMIONS, [D83].

Let Ψ and ρ_Ψ be as in (2.6). Then

$$\langle \Psi | \sum_{i=1}^N |p_i| | \Psi \rangle \geq 1.63q^{-1/3} \int_{\mathbf{R}^3} \rho_\Psi^{4/3}(x) dx . \quad (3.5)$$

A generalization of this, of importance if we wish to include a mass, is

$$\langle \Psi | \sum_{i=1}^N [\sqrt{p_i^2 + m^2} - m] | \Psi \rangle \geq \frac{3}{8} m^4 C \int_{\mathbf{R}^3} g \left((\rho_{\Psi}(x)/C)^{1/3} m^{-1} \right) dx, \quad (3.6)$$

with $C = 0.163q$ (*sic*) and with

$$g(t) := t(1 + t^2)^{1/2}(1 + 2t^2) - \frac{8}{3}t^3 - \ln \left[t + (1 + t^2)^{1/2} \right]. \quad (3.7)$$

GENERAL KINETIC ENERGY BOUND, [C84], p.454, (and [HO77] for the nonrelativistic case). The following bound follows from a judicious application of Schwarz's inequality.

$$\langle \Psi | \sum_{i=1}^N |p_i| | \Psi \rangle \geq (\sqrt{\rho_{\Psi}}, |p| \sqrt{\rho_{\Psi}}). \quad (3.8)$$

This bound holds irrespective of the symmetry type of the wave function.

EXCHANGE AND CORRELATION INEQUALITY, [LO81]. If Ψ is a normalized N -particle wave function there is a lower bound on the interparticle Coulomb repulsion in terms of its density:

$$\langle \Psi | \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} | \Psi \rangle \geq D(\rho_{\Psi}, \rho_{\Psi}) - 1.68 \int_{\mathbf{R}^3} \rho_{\Psi}^{4/3}(x) dx. \quad (3.9)$$

(Once again, the antisymmetry of Ψ plays no role in this inequality.)

ELECTROSTATIC INEQUALITY, [LY88], p.196. First, we define a function, Φ on $\mathbf{R}^3$ with the aid of the Voronoi cells mentioned above. In the cell Γ_j , Φ equals the electrostatic potential generated by all the nuclei *except* for the nucleus situated in Γ_j itself, i.e., for x in Γ_j

$$\Phi(x) := Z \sum_{\substack{i=1 \\ i \neq j}}^K |x - R_i|^{-1}. \quad (3.10)$$

If ν is any bounded Borel measure on $\mathbf{R}^3$ (not necessarily positive) then

$$\frac{1}{2} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} |x - y|^{-1} d\nu(x) d\nu(y) - \int_{\mathbf{R}^3} \Phi(x) d\nu(x) + U \geq \frac{1}{8} Z^2 \sum_{j=1}^K D_j^{-1}. \quad (3.11)$$

4. Proofs of Theorems 1 and 2

To prove Theorem 1 we take $\beta = \pi Z\alpha/2$ (if $\beta > \pi Z\alpha/2$ we simply throw away the excess positive quantity). Using (3.2) with f replaced by $\sqrt{\rho}$, we have that

$$\mathcal{E}(\rho) \geq \mathcal{E}_1(\rho) + \alpha \mathcal{E}_2(\rho), \quad (4.1)$$

where, by adding and subtracting a term $\int \Phi \rho$, with $\Phi(x)$ as in (3.10),

$$\mathcal{E}_1(\rho) := \frac{3}{4} \gamma \int_{\mathbf{R}^3} \rho^{4/3}(x) dx - \alpha \int_{\mathbf{R}^3} W(x) \rho(x) dx + \alpha \int_{\mathbf{R}^3} \Phi(x) \rho(x) dx \quad (4.2)$$

and

$$\mathcal{E}_2(\rho) := D(\rho, \rho) - \int_{\mathbf{R}^3} \Phi(x) \rho(x) dx + U. \quad (4.3)$$

The function $W(x)$ is defined as follows: In the Voronoi cell Γ_j it is given by

$$W(x) := \Phi(x) + \begin{cases} Z|x - R_j|^{-1}, & \text{if } |x - R_j| > D_j \\ (\pi Z/2) D_j^{-1} Y(|x - R_j|/D_j), & \text{if } |x - R_j| \leq D_j. \end{cases} \quad (4.4)$$

Note that while the terms $\pm \int \Phi \rho$ that appear in (4.2), (4.3) are merely 'strategic', the presence of the term $\Phi(x)$ in (4.4) is properly part of the potential energy of the electron and is not arbitrary. Actually, this strategic decomposition of $\mathcal{E}(\rho)$ is the one used in the easy part of Fenchel's duality theorem (see [R70], p. 327). This duality principle was used in connection with Thomas-Fermi theory by Firsov [F57] (see also [L81]); the full blown duality theory is not needed for our purposes, so we omit it.

We can now seek lower bounds for $\mathcal{E}_1(\rho)$ and $\mathcal{E}_2(\rho)$ separately. Using Hölder's inequality, for example, one easily concludes that the absolute minimum of $\mathcal{E}_1(\rho)$ is

$$\begin{aligned} \mathcal{E}_1(\rho) &\geq -\frac{\alpha^4}{4\gamma^3} \int_{\mathbf{R}^3} [W(x) - \Phi(x)]_+^4 dx \\ &= -\frac{(\alpha Z)^4}{4\gamma^3} \sum_{j=1}^K \left(\frac{\pi}{2}\right)^4 \int_{B_j} D_j^{-4} Y(|x - R_j|/D_j)^4 dx + \int_{\Gamma_j \setminus B_j} |x - R_j|^{-4} dx \end{aligned} \quad (4.5)$$

$$\geq -\frac{(\alpha Z)^4}{4\gamma^3} \left\{ \left(\frac{\pi}{2}\right)^4 (4\pi) \int_0^1 Y(r)^4 r^2 dr + 3\pi \right\} \sum_{j=1}^K D_j^{-1} \quad (4.6)$$

$$> -\frac{(\alpha Z)^4}{4\gamma^3} \left\{ 7.6245 \left(\frac{\pi}{2}\right)^4 + 3\pi \right\} \sum_{j=1}^K D_j^{-1}. \quad (4.7)$$

The last formula uses (3.4). The second integral in (4.5) is evaluated in (4.6) as $3\pi/D_j$, and the explanation is the following: If we integrate $|x - R_j|^{-4}$ over the exterior of B_j we would obtain $4\pi/D_j$ as the result. However, we know that the Voronoi cell Γ_j lies on

one side of the mid-plane defined by the nearest neighbor nucleus. This means that the integral over $\Gamma_j \setminus B_j$ is bounded above by the quantity

$$\frac{4\pi}{D_j} - \frac{1}{D_j} \int_1^\infty dz \int_0^\infty \frac{2\pi r}{(r^2 + z^2)^2} dr = \frac{3\pi}{D_j} .$$

The $\mathcal{E}_2$ term can be bounded using (3.11) with $d\nu(x) = \rho(x)dx$. Thus,

$$\mathcal{E}_2(\rho) \geq \frac{Z^2}{8} \sum_{j=1}^K \frac{1}{D_j} . \quad (4.8)$$

Combining (4.1), (4.7) and (4.8) we have proved Theorem 1. ■

Theorem 2 is proved by splitting the relativistic kinetic energy $|p|$ into $\beta|p|$ and $(1 - \beta)|p|$, with the choice $\beta = \pi Z\alpha/2$. The inequalities (3.5), (3.8) and (3.9) immediately give us Theorem 2. ■

5. Inclusion of Mass and Magnetic Fields

INCLUSION OF MASS. We replace $|p|$ by $\sqrt{p^2 + m^2} - m$ and, in the corresponding TFW theory, we replace the right side of (3.5) by the right side of (3.6). It is not easy to carry out the rest of the program in closed form with this more complicated function, however. Moreover, it unfortunately gives a slightly worse constant than before, even when we set $m = 0$; instead of $1.63q^{-1/3}$ in (3.5) we now have $C^{-1/3} \approx 1.37q^{-1/3}$. The new energy will not be positive in the stability regime, as we had before. Instead, it will be a negative constant times N . This new value for the energy is in accord with stability of matter and represents the binding energy of the electron-nuclear system.

Another way to deal with the mass is to observe, simply, that $\sqrt{p^2 + m^2} - m > |p| - m$, the effect of which is to add a term $-Nm$ to the energy estimate. This term satisfies the criterion for stability, but it has the defect that is huge in real-world terms, for it equals the rest energy of the electron.

INCLUSION OF MAGNETIC FIELD. Theorem 2, with a magnetic field included, is a consequence of the following two inequalities (proved below) which replace (3.6) and (3.8):

$$\langle \Psi | \sum_{i=1}^N [\sqrt{(p_i + A(x_i))^2 + m^2} - m] | \Psi \rangle \geq \frac{3}{8} m^4 C \int_{\mathbf{R}^3} g \left((\rho_\Psi(x)/C)^{1/3} m^{-1} \right) dx , \quad (5.1)$$

and

$$\langle \Psi | \sum_{i=1}^N |p_i + A(x_i)| | \Psi \rangle \geq (\sqrt{\rho_\Psi} , |p| \sqrt{\rho_\Psi}) . \quad (5.2)$$

As in (3.8), inequality (5.2) holds irrespective of the symmetry type of Ψ .

To define $\sqrt{|p+A|^2+m^2}$, note that if $A \in L^2_{loc}(\mathbf{R}^3; \mathbf{R}^3)$, then $f \mapsto \|(p+A)f\|_2^2$ is a closed quadratic form with $C_0^\infty(\mathbf{R}^3)$ being a form core [K78], [S79-1], [LS81]. Thus it defines a selfadjoint operator and it is then possible to define $\sqrt{|p+A|^2+m^2}$ via the spectral calculus.

The *diamagnetic inequality for the heat kernel* [S79-2] is the pointwise inequality

$$| \exp[-t(p+A)^2] f(x) | \leq \exp[-tp^2] |f|(x) . \quad (5.3)$$

Using the formula

$$e^{-|a|} = \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-t-a^2/4t} \frac{dt}{\sqrt{t}} , \quad (5.4)$$

which holds for any real number a (and hence for any selfadjoint operator), we obtain the diamagnetic inequality for the 'relativistic heat kernel'

$$| \exp[-t\sqrt{(p+A)^2+m^2}] f(x) | \leq \exp[-t\sqrt{p^2+m^2}] |f|(x) . \quad (5.5)$$

By using (5.5), and following the proof of (3.6) in [D83] step by step, we obtain (5.1). Likewise, (5.5) and the formula

$$(f, \sqrt{(p+A)^2+m^2} f) = \lim_{t \rightarrow 0} \frac{1}{t} \left\{ (f, f) - (f, \exp[-t\sqrt{(p+A)^2+m^2}] f) \right\} , \quad (5.6)$$

yield

$$(f, |p+A| f) \geq (|f|, |p| |f|) . \quad (5.7)$$

To prove (5.2) we apply (5.7) to the function $|\Psi|$ and then use (3.8).

References

- [C84] Conlon, J.G., *The ground state energy of a classical gas*, *Commun. Math. Phys.* **94**, 439-458 (1984).
- [D83] Daubechies, I., *An uncertainty principle for fermions with generalized kinetic energy*, *Commun. Math. Phys.* **90**, 511-520 (1983).
- [F57] Firsov, O.B., *Calculation of the interaction potential of atoms for small nuclear separations*, *Sov. Phys. JETP* **5**, 1192-1196 (1957).
- [HO77] Hoffmann-Ostenhof, M. and Hoffman-Ostenhof, T., *Schrödinger inequalities and asymptotic behavior of the electronic density of atoms and molecules*, *Phys. Rev. A* **16**, 1782-1785 (1977).

-
- [K78] Kato, T., *Remarks on Schrödinger operators with vector potentials*, Int. Eq. Operator Theory **1**, 103-113 (1978).
 - [LS81] Leinfelder, H., Simader, C., *Schrödinger operators with singular magnetic vector potentials*, Math. Z. **176**, 1-19 (1981).
 - [L81] Lieb, E.H. *Thomas-Fermi and related theories of atoms and molecules*, Rev. Mod. Phys. **53** 603-641 (1981). Errata, *ibid* **54**, 311 (1982).
 - [LLS95] Lieb, E.H., Loss, M. and Solovej, J.P., *Stability of Matter in Magnetic Fields*, Phys. Rev. Lett. **75**, 985-989 (1995).
 - [LO81] Lieb, E.H. and Oxford, S., *Improved lower bound on the indirect Coulomb energy*, Int. J. Quant. Chem. **19**, 427-439 (1981).
 - [LY88] Lieb, E.H. and Yau, H-T., *The stability and instability of relativistic matter*, Commun. Math. Phys. **118**, 177-213 (1988).
 - [LT75] Lieb, E.H., and Thirring, W.E., *Bound for the kinetic energy of fermions which proves the stability of matter*, Phys. Rev. Lett. **35**, 687-689 (1975). Erratum, *ibid*, 1116.
 - [R70] Rockafellar, R.T., *Convex Analysis*, Princeton University Press (1970).
 - [S79-1] Simon, B., *Maximal and minimal Schrödinger forms*, J. Opt. Theory **1**, 37-47 (1979).
 - [S79-2] Simon, B., *Kato's inequality and the comparison of semigroups*, J. Funct. Anal. **32**, 97-101 (1979).

Stability and Instability of Relativistic Electrons in Classical Electromagnetic Fields

Elliott H. Lieb,¹ Heinz Siedentop,² and Jan Philip Solovej³

Received October 21, 1996

The stability of matter composed of electrons and static nuclei is investigated for a relativistic dynamics for the electrons given by a suitably projected Dirac operator and with Coulomb interactions. In addition there is an arbitrary classical magnetic field of finite energy. Despite the previously known facts that ordinary nonrelativistic matter with magnetic fields, or relativistic matter without magnetic fields, is already unstable when α , the fine structure constant, is too large, it is noteworthy that the combination of the two is still stable *provided* the projection onto the positive energy states of the Dirac operator, which *defines* the electron, is chosen properly. A good choice is to include the magnetic field in the definition. A bad choice, which always leads to instability, is the usual one in which the positive energy states are defined by the free Dirac operator. Both assertions are proved here.

KEY WORDS: Stability of matter; Schrödinger operators; magnetic fields; relativistic; Dirac operator; instability of matter.

1. INTRODUCTION

The stability of matter concerns the many-electron and many-nucleus quantum mechanical problem and the question whether the ground state energy is finite (stability of the first kind). If so, is it bounded below by a

¹ Department of Mathematics and Physics, Princeton University, Princeton, New Jersey 08544-0708.

² Mathematik, Universität Regensburg, D-93040 Regensburg, Germany.

³ Institut for matematiske fag, Aarhus Universitet, Ny Munkegade, DK-8000 Aarhus C, Denmark. Current address: Department of Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.

©1996 by the authors. Reproduction of this article, in its entirety, by any means is permitted for noncommercial purposes.

This paper is dedicated to Bernard Jancovici on the occasion of his 65th birthday.

constant (which is independent of the position of the nuclei) times the number of particles (stability of the second kind)? The linear lower bound is important for thermodynamics, which will not exist in the usual way without it.

The first positive resolution of this problem for the nonrelativistic Schrödinger equation was given by Dyson and Lenard^(7,8) and approached differently by Federbush.⁽¹⁰⁾ The constant, i.e., the energy per particle, was considerably improved by Lieb and Thirring in refs. 21 and 22. Following that, the stability of a relativistic version of the Schrödinger equation (in which $\mathbf{p}^2$ is replaced by $\sqrt{\mathbf{p}^2 + m^2}$) was proved by Conlon⁽⁵⁾ and later improved by Lieb and Yau⁽²³⁾ who showed that matter is stable in this model if and only if the fine structure constant α is small enough and if $Z\alpha \leq 2/\pi$. (See ref. 23 for a historical account up to 1995.) A recent result of Lieb, Loss, and Siedentop that we shall use is in ref. 19 and is discussed in Section 3.

In these works the nuclei are fixed in space because they are very massive and because we know that the nuclear motion is largely irrelevant for understanding matter. In other words, if nuclear motion were the only thing that prevented the instability of matter then the world would look very different from what it does. We continue this practice here.

There is, however, a more important quantity that requires some attention, namely magnetic fields. It was noted that the action of such fields on the translational degrees of freedom of the electrons $\mathbf{p} \rightarrow \mathbf{p} + e\mathbf{A}$, can lower the energy only by an inconsequential amount. This is a kind of diamagnetic inequality. On the other hand, spin-magnetic field interaction (in which $(\mathbf{p} + e\mathbf{A})^2$ is replaced by the Pauli operator $[\boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})]^2 = (\mathbf{p} + e\mathbf{A})^2 + e\boldsymbol{\sigma} \cdot \mathbf{B}$) can cause instability. The energy is then unbounded below if arbitrarily large fields are allowed, but this is so only because the energy of the magnetic field has not been taken into account. Does the field energy, $(8\pi)^{-1} \int B^2$, insure stability? This question was raised for the non-relativistic case in ref. 13 and finally settled in a satisfactory manner in ref. 20 (see also Bugliaro *et al.*⁽⁴⁾ and Fefferman⁽¹¹⁾). The upshot of this investigation is that stability (of both first and second kinds) requires a bound on both α and on $Z\alpha^2$.

Other related results are the stability of non-relativistic matter with a second quantized, ultra-violet cut-off photon field (Fröhlich *et al.*⁽¹²⁾).

Both the passage to relativistic kinematics (which, in quantum mechanics, means that both the kinetic energy and the Coulomb potential scale with length in the same way, namely like an inverse length) and the introduction of the nonrelativistic Pauli operator require a bound on α and on Z for stability. The combination of the two might be expected to lead to disaster. We find, however, that it does not necessarily do so!

Our main result is that matter is indeed stable with a suitably defined relativistic kinematics. This is shown in Section 3.

The proper way to introduce relativistic kinematics for spin-1/2 particles is via the Dirac operator, but this is unbounded below. A resolution of this problem, due to Dirac, is to permit the electrons to live only in the positive energy subspace of the Dirac operator. This idea was further pursued by Brown and Ravenhall⁽³⁾ (see also Bethe and Salpeter in their Handbuch article⁽¹⁾) to give a quantitative description of real atoms.

There are, however, other Dirac operators (which include electromagnetic potentials) whose positive subspace can be used to define the space in which the electrons can live. (To avoid confusion, let us note that the Hamiltonian is formally always the same and includes whatever fields happen to be present. The only point to be resolved is what part of the one-particle Hilbert space is allowed for electrons.) The review articles of Sucher^(24, 25, 26) can be consulted here. These choices have also been used in quantum chemistry and other practical calculations, see, e.g., refs. 14 and 15.

All of these choices have in common that there is no creation of electron-positron pairs explaining the name “no-pair Hamiltonian” for the resulting energy operator. (Note that we could also treat positrons or a combination of electrons and positrons interacting by Coulomb forces in a similar way.)

There are three obvious choices to consider. One is the free Dirac operator. This always leads to instability of the first kind when a magnetic field is added unless the particle number is held to some small value (see Section 4). Note also that this choice leads to a non-gauge invariant model: multiplication of a state with the factor $\exp(i\phi(\mathbf{r}))$ for a non-constant gauge is not allowed, since it leads out of the positive spectral subspace.

Remarkably, the Dirac operator that includes the magnetic field always gives stability, if Z and α are not too large, as in the two cases (relativistic without magnetic field and nonrelativistic with magnetic field) mentioned above (see Section 3). This model is gauge invariant.

The third choice which, indeed, is sometimes used, is to include both the one-body attractive electric potential of the nuclei and the magnetic field in the definition of the Dirac operator that defines the positive subspace. If this is done then the question of stability is immediately solved because the remaining terms in the Hamiltonian are positive, and hence the total energy is ipso facto positive. This choice, which is important but trivial in the context of this present inquiry, will not be mentioned further.

Oddly, the instability proof given in Subsection 4.2 is much more complicated than the stability proof (Section 3). This reverses the usual situation.

A preliminary version of this paper appeared in ref. 27; the present version is to be regarded as the original one (as stated in ref. 27) and contains several significant corrections to the preliminary text in ref. 27. In particular, the proof of Theorem 2 and the first half of the proof of Theorem 1 have been corrected and simplified. A summary of this work appears in ref. 28.

2. BASIC DEFINITIONS

The energy of N relativistic electrons in the field of K nuclei with atomic numbers $Z_1, \dots, Z_K \in \mathbb{R}_+$ located at $\mathbf{R}_1, \dots, \mathbf{R}_K \in \mathbb{R}^3$ which are pairwise different in a magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ in the state Ψ is—following the ideas of Brown and Ravenhall⁽³⁾—

$$\mathcal{E}[\Psi, \Psi] := \left(\Psi, \left(\sum_{v=1}^N D_v(\mathbf{A}) + \alpha V_c \right) \Psi \right) + \frac{1}{8\pi} \int_{\mathbb{R}^3} B(\mathbf{r})^2 d\mathbf{r}. \quad (1)$$

Here $D_v(\mathbf{A}) := \boldsymbol{\alpha} \cdot (-i \nabla_v + e \mathbf{A}(\mathbf{r}_v)) + m\beta$ is the Dirac operator with vector potential $\mathbf{A}$. Furthermore,

$$V_c := - \sum_{v=1}^N \sum_{\kappa=1}^K \frac{Z_\kappa}{|\mathbf{r}_v - \mathbf{R}_\kappa|} + \sum_{\mu, v=1}^N \sum_{\mu < v} \frac{1}{|\mathbf{r}_\mu - \mathbf{r}_v|} + \sum_{\kappa, \lambda=1}^K \sum_{\kappa < \lambda} \frac{Z_\kappa Z_\lambda}{|\mathbf{R}_\kappa - \mathbf{R}_\lambda|} \quad (2)$$

is the Coulomb interaction between the particles, and $B(\mathbf{r}) := |\nabla \times \mathbf{A}(\mathbf{r})|$ is the modulus of the magnetic field. Planck's constant divided by 2π and the velocity of light, are taken to be one in suitable units. The fine structure constant α equals e^2 , where $-e$ is the electron charge. Experimentally, α is about $1/137.037$. The mass of the electron is denoted by m . The 4×4 matrices $\boldsymbol{\alpha}$ and β are the four Dirac matrices in standard representation, namely

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and

$$\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Finally, the state Ψ should have finite kinetic energy, i.e., it should be in the Sobolev space $H^{1/2}[(\mathbb{R}^3 \times \{1, 2, 3, 4\})^N]$, and should also be in the electronic Hilbert space of antisymmetric spinors

$$\mathcal{H}_{N, \mathcal{A}} := \bigwedge_{v=1}^N \mathcal{H}_+, \quad (3)$$

where $\mathcal{H}_+$ is the positive spectral subspace of the Dirac operator $D(\mathcal{A})$ and where $\mathcal{A}$ is some vector potential to be chosen later. The vector potential $\mathcal{A}$ serves to define the positive subspace. The restriction of $\mathcal{E}$ to $\mathcal{H}_{N, \mathcal{A}}$ will be denoted $\mathcal{E}_{\mathcal{A}}$. Two choices will be considered here. One is $\mathcal{A} = \mathbf{0}$, in which case we are talking about the free Dirac operator. This choice, or model, goes back to Brown and Ravenhall.⁽³⁾ As we shall see in Section 4, the resulting energy functional—apart from being not gauge invariant—is not bounded from below. A natural modification of the model, namely to take $\mathcal{A} := \mathbf{A}$ is not only gauge invariant, but will also turn out to be stable of the second kind (see Section 3).

The quantity of interest is the lowest possible energy

$$E_{N, K} := \inf \mathcal{E}_{\mathcal{A}}$$

where the infimum is taken over all allowed, normalized states Ψ , all allowed vector potentials $\mathbf{A}$, and over all pairwise different nuclear positions $\mathbf{R}_1, \dots, \mathbf{R}_K$.

In the case of a single nucleus without a magnetic field, the energy form $\mathcal{E}_0$ was shown in ref. 9 to be bounded from below, if and only if $\alpha Z \leq \alpha Z_C := 2/(\pi/2 + 2/\pi) > 2/\pi$, which corresponds to $Z_C \approx 124$. We will not be able to reach this value in the general case of many nuclei and when the electron state space is not determined by the free Dirac operator. The reason is that special techniques were used in ref. 9 to handle the one-nucleus case; these techniques took advantage of the weakening of the Coulomb singularities caused by the fact that states in $\mathcal{H}_+$ cannot be localized in space arbitrarily sharply. Unfortunately, we do not know how to implement this observation with magnetic fields and many nuclei.

3. STABILITY WITH THE MODIFIED PROJECTOR

Our proof of the stability of matter when the vector potential $\mathbf{A}$ is included in the definition of the positive energy electron states will depend essentially on three inequalities:

BKS Inequality. For any self-adjoint operator X , the negative (positive) part, $X_{\mp}$ is defined to be $(|X| \mp X)/2$. Given two non-negative

self-adjoint operators C and D such that $(C^2 - D^2)_-^{1/2}$ is trace class, we have the trace inequality

$$\mathrm{tr}(C - D)_- \leq \mathrm{tr}(C^2 - D^2)_-^{1/2}. \quad (4)$$

This is a special case of a more general inequality of Birman, Koplienko, and Solomyak;⁽²⁾ in particular, the number 2 in (4) can be replaced by any $p > 1$. A proof for the special case of the inequality needed here is given in Appendix A.

Stability of Relativistic Matter. On $\bigwedge_{v=1}^N (H^{1/2}(\mathbb{R}^3) \otimes \mathbb{C}^q)$, the fermionic Hilbert space, we have

$$\sum_{v=1}^N |-i \nabla_v - \mathcal{A}| + \tilde{\alpha} V_c \geq 0, \quad (5)$$

(where $|\dots|$ means $\sqrt{(\dots)^2}$) for all vector fields $\mathcal{A}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ with, e.g., square integrable gradients, if

$$1/\tilde{\alpha} \geq 1/\tilde{\alpha}_c := (\pi/2) Z + 2.2159q^{1/3}Z^{2/3} + 1.0307q^{1/3} \quad (6)$$

and $Z_1, \dots, Z_K \leq Z$.

We wish to use this inequality for 4-component spinors, i.e., $q=4$. However, we are interested in the subspace $\mathcal{H}_{N, \mathcal{A}}$ in which the particles are restricted to the positive energy subspace of the Dirac operator, $D(\mathcal{A})$. Although $q=4$, the “effective” q is really 2, and the analysis in Appendix B is our justification for this. The only thing that really counts in deriving (5) is the bound on the reduced one-body density matrix γ mentioned in Appendix B.

The stability of the relativistic Hamiltonian (5) was first shown by Conlon⁽⁵⁾ for $\mathcal{A} = \mathbf{0}$. The best currently available constants with $\mathcal{A} = \mathbf{0}$ are in ref. 23 while (6), which is taken from ref. 19, is the best known result for general $\mathcal{A}$.

Semi-Classical Bound. Given a positive constant μ , a real vector field $\mathcal{A}$ with, e.g., square integrable gradients, and a real-valued function $\varphi \in L^2(\mathbb{R}^3)$ the inequality

$$\mathrm{tr}[(-i\mu \nabla - \mathcal{A})^2 - \varphi]_-^{1/2} \leq \frac{L_{1/2, 3}}{\mu^3} \int_{\mathbb{R}^3} \varphi^2_+ \quad (7)$$

holds, which is a special case of the Lieb-Thirring inequality (see refs. 22 and 17]). It is known that $L_{1/2, 3} \leq 0.06003$. The left side of (7) is simply $\sum_j |\lambda_j|^{1/2}$, where the λ_j are the negative eigenvalues of the operator $[\dots]$.

As an illustration of the usefulness of the trace estimate (4), let us combine it with the Lieb–Thirring inequality (7) (or any other Lieb–Thirring inequality for that matter) to derive some previously known inequalities. The constants obtainable in this way are comparable with the best ones known so far:

Daubechies Inequality. We begin with a “relativistic” inequality that was first proven by Daubechies.⁽⁶⁾ By replacing φ by φ^2 in (7), we get using (4)

$$\mathrm{tr}(|-i\nabla - \mathcal{A}| - \varphi)_- \leq L_{1/2,3} \int_{\mathbb{R}^3} \varphi_+^4. \quad (8)$$

The constant 0.06003 obtained here should be compared with the number 0.0258 in ref. 6.

Non-Relativistic Magnetic Stability. A non-relativistic analogue of our main problem is to bound the form

$$\tilde{\mathcal{E}} := \left(\Psi, \left(\sum_{\nu=1}^N P_{\nu}(\mathbf{A}) + \alpha V_c \right) \Psi \right) + \frac{1}{8\pi} \int_{\mathbb{R}^3} B(\mathbf{r})^2 d\mathbf{r} \quad (9)$$

which was treated in ref. 20. Here $P_{\nu}(\mathbf{A}) := [\boldsymbol{\sigma} \cdot (-i\nabla_{\nu} + e\mathbf{A}(\mathbf{r}_{\nu}))]^2$ is the Pauli operator with vector potential $\mathbf{A}$.

First, we note that $x^2 \geq +\lambda|x| - \lambda^2/4$ holds. A constant in the energy form, however, is irrelevant for checking on stability of the second kind. Using (5) it is then enough to show the positivity of

$$-\mathrm{tr}(\lambda P(\mathbf{A})^{1/2} - \kappa |-i\nabla + e\mathbf{A}|)_- + \frac{1}{8\pi} \int_{\mathbb{R}^3} B(\mathbf{r})^2 d\mathbf{r} \quad (10)$$

where we have set $\kappa := \alpha/\tilde{\alpha}_c$. The trace in this and the next expression are over $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$. Using the BKS inequality gives the lower bound

$$-\mathrm{tr}[(\lambda^2 - \kappa^2) |-i\nabla + e\mathbf{A}|^2] - e\lambda^2 B(\mathbf{r})]_-^{1/2} + \frac{1}{8\pi} \int_{\mathbb{R}^3} B(\mathbf{r})^2 d\mathbf{r}.$$

Applying the Lieb–Thirring inequality (7) yields the following sufficient condition for stability (recall that $\alpha = e^2$ and that there are two spin states)

$$2L_{1/2,3} \frac{\lambda^4 \alpha}{(\lambda^2 - \kappa^2)^{3/2}} \leq \frac{1}{8\pi}.$$

Optimizing in λ gives

$$2L_{1/2, 3} \frac{16\alpha^2}{3^{3/2}\tilde{\alpha}_c} \leq \frac{1}{8\pi},$$

which gives for the physical values $\alpha \approx 1/137.037$ and $q=2$ a range of stability up to $Z \leq 1096$, which is to be compared with $Z \leq 1050$ in ref. 20.

We turn now to our main result.

Theorem 1. Let $Z_1, \dots, Z_K \leq Z < 2/(\pi\alpha)$ and let $\alpha \leq \alpha_c$ where α_c is the unique solution of the equation

$$(16\pi L_{1/2, 3} \alpha_c)^{2/3} = 1 - \alpha_c^2 / \tilde{\alpha}_c^2,$$

with $\tilde{\alpha}_c := [(\pi/2) Z + 2.2159 \cdot 2^{1/3} Z^{2/3} + 1.0307 \cdot 2^{1/3}]^{-1}$ as in (6). Then $\mathcal{E}_A$ is non-negative.

Numerically, this gives

$$Z \leq 56$$

when evaluated with the experimental value $\alpha \approx 1/137.037$ for the fine structure constant. Alternatively, considering hydrogen only, i.e., $Z=1$, we obtain the upper bound

$$\alpha \leq 1/8.139$$

for the fine structure constant. It is a challenge to improve this result so that it covers all physical nuclear charges and the physical value of the fine structure constant, as was done for $K=N=1$ and $A=0$ in ref. 9.

It is easy to prove (we do not do so here) that, as expected, $\mathcal{E}_A/(\Psi, \Psi) \geq m - O(\alpha^2 Z^2)$ for small α and Z .

Proof. The first step in our proof is to utilize (5) to replace V_c by the one-body operator $(-1/\tilde{\alpha}_c) \sum_{v=1}^N |-i \nabla_v + e \mathbf{A}|$, where $\tilde{\alpha}_c$ is given by (6) with $q=2$, as we explained just after (6). (The idea of using the relativistic stability result (5) to bound the Coulomb potential by a one-body operator first appears in ref. 20.) Our energy $\mathcal{E}_A$ is now bounded below by

$$\mathcal{E}'(\Psi) := \left(\Psi, \sum_{v=1}^N (D_v(\mathbf{A}) - \kappa |-i \nabla_v + e \mathbf{A}(\mathbf{r}_v)|) \Psi \right) + \frac{1}{8\pi} \int_{\mathbb{R}^3} B(\mathbf{r})^2 d\mathbf{r}, \quad (11)$$

where $\kappa := \alpha/\tilde{\alpha}_c$.

The first term on the right side of (11) is bounded below by the sum of the negative eigenvalues, $-\text{tr } h_-$, of the one-body operator

$$h := A^+ (D(\mathbf{A}) - \kappa |-i \nabla + e \mathbf{A}(\mathbf{r})|) A^+,$$

where A^+ is the projector onto the positive spectral subspace of $D(\mathbf{A})$.

Let us define

$$S := |D(\mathbf{A})| - \kappa |-i \nabla + e \mathbf{A}(\mathbf{r})|,$$

whence $h = A_+ S A_+$, because $A_+ D(\mathbf{A}) A_+ = A_+ |D(\mathbf{A})| A_+$. We note that for any two self-adjoint operators X and ρ with $X \geq 0$ and $0 \leq \rho \leq 1$, $\text{tr } X \geq \text{tr } \rho X$. With ρ taken to be the projector onto the negative spectral subspace of h we then have that

$$\begin{aligned} \text{tr } h_- &= -\text{tr } \rho h = -\text{tr } \rho A_+ S A_+ \\ &= \text{tr } \rho A_+ S_- A_+ - \text{tr } \rho A_+ S_+ A_+ \\ &\leq \text{tr } A_+ S_- A_+. \end{aligned} \tag{12}$$

We introduce the 4×4 unitary

$$U = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and note that $U^{-1} D(\mathbf{A}) U = -D(\mathbf{A})$. Therefore, $U^{-1} A_+ U = A_-$.

It follows from the spectral theorem that for any self-adjoint X , unitary U , and function F

$$F(U^{-1} X U) = U^{-1} F(X) U.$$

With $F(t) = |t|$, we then have that $U^{-1} |D(\mathbf{A})| U = |D(\mathbf{A})|$, and hence $U^{-1} S U = S$.

Therefore, since $U^{-1} A_+ U = A_-$, and with $F(t) = \frac{1}{2}(|t| - t) = t_-$, we have that $U^{-1} S_- U = S_-$ and

$$\text{tr } A_+ S_- A_+ = \text{tr } A_- S_- A_-.$$

Hence, using (12),

$$\text{tr } h_- \leq \frac{1}{2} \text{tr}(A_+ S_- A_+) + \frac{1}{2} \text{tr } A_- S_- A_- = \frac{1}{2} \text{tr } S_-.$$

(Note: much of the preceding discussion was needed only to get the factor $1/2$ here. This factor improves our final constants for stability.)

Next, we use the BKS inequality (4) to bound $\text{tr } S_-$ as follows:

$$\text{tr } h_- \leq \frac{1}{2} \text{tr } S_- \leq \frac{1}{2} \text{tr} [D(\mathbf{A})^2 - \kappa^2 |-i \nabla + e \mathbf{A}(\mathbf{r})|^2]_-^{1/2}. \tag{13}$$

However, $D(\mathbf{A})^2 = \begin{pmatrix} Y & 0 \\ 0 & Y \end{pmatrix}$ with $Y = P(\mathbf{A}) + m^2$, and where

$$P(\mathbf{A}) = [\boldsymbol{\sigma} \cdot (-i \nabla + e \mathbf{A})]^2 = |-i \nabla + e \mathbf{A}(\mathbf{r})|^2 + e \boldsymbol{\sigma} \cdot \mathbf{B}(\mathbf{r})$$

is the Pauli operator.

Since $X \mapsto \text{tr } X^{1/2}$ is operator monotone decreasing, we see that our lower bound for the energy is monotone increasing in m , and thus it suffices to prove the positivity of $\mathcal{H}_{N, \mathbf{A}}$ in the massless case. The key observation is that our lower bound involves only $\text{tr } S_-$ in the entire one-body space, not the positive energy subspace. The energy would not be obviously monotone in m if we had to restrict functions to the positive subspace, since changing m would also entail changing the space. This problem does not arise in the absence of the positive subspace constraint.

Because of the “diagonal” structure of the operator S , we can drop the factor $1/2$ by replacing the trace on $L^2(\mathbb{R}^3) \otimes \mathbb{C}^4$ by the trace on $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$. This yields

$$\begin{aligned} \mathcal{E}_{\mathbf{A}}[\Psi, \Psi] &\geq -\text{tr}[P(\mathbf{A}) - \kappa^2(-i \nabla + e \mathbf{A})^2]_-^{1/2} + \frac{1}{8\pi} \int_{\mathbb{R}^3} B(\mathbf{r})^2 d\mathbf{r} \\ &\geq -2 \text{tr}[(1 - \kappa^2)(-i \nabla + e \mathbf{A})^2 - eB]_-^{1/2} + \frac{1}{8\pi} \int_{\mathbb{R}^3} B(\mathbf{r})^2 d\mathbf{r}. \end{aligned} \quad (14)$$

We regard the operator in the second line as acting on functions (of one component only) instead of spinors, which accounts for the factor two (and not one). Finally we apply the Lieb–Thirring inequality (7) to the right hand side yielding (recall that $e^2 = \alpha$)

$$\mathcal{E}_{\mathbf{A}}[\Psi, \Psi] \geq [-2L_{1/2, 3}\alpha(1 - \alpha^2/\tilde{\alpha}_c^2)^{-3/2} + 1/(8\pi)] \int_{\mathbb{R}^3} B^2(\mathbf{r})^2 d\mathbf{r}.$$

Thus we need

$$(16\pi L_{1/2, 3}\alpha)^{2/3} \leq 1 - \alpha^2/\tilde{\alpha}_c^2. \quad (15)$$

Since the right hand side of this inequality is monotone decreasing in α for positive α , while the left hand side is monotone increasing, there is a unique α_c for which equality holds in (15). Inserting the value (6) with $q=2$ for $\tilde{\alpha}_c$ yields—together with the second requirement on $Z_1, \dots, Z_K$ in the relativistic bound—the claimed stability criterion.

4. INSTABILITY WITH THE FREE DIRAC OPERATOR

In this section we shall discuss the Brown–Ravenhall model.⁽³⁾ That is to say we consider the energy expression (1) with zero vector potential in the definition of the allowed electronic states (3), i.e., we take only $\Psi \in \mathcal{H}_{N,0} = \bigwedge^N \mathcal{H}_+$, where $\mathcal{H}_+$ is the positive spectral subspace of the operator $-\alpha \cdot i \nabla + m\beta$. We shall prove that there is no stability in this model by showing that for any (sufficiently large) particle number, N , and any $\alpha > 0$ the energy is unbounded below. In other words, “stability of the first kind” is violated. It is nevertheless true, however, that for any choice of particle numbers and nuclear charges there is always a sufficiently small, nonzero α such that the energy is bounded below by zero.

Since the positive spectral subspace $\mathcal{H}_+$ for the free Dirac operator is not invariant under gauge transformations we see that this Brown–Ravenhall model is not gauge invariant. (The previous, modified model discussed in Section 3 is not only stable, it is also gauge invariant.) More precisely, the energy spectrum depends not only on $\mathbf{B}$ but in fact on the full gauge potential $\mathbf{A}$. The Brown–Ravenhall model is therefore physically meaningfully defined only if we make a fixed choice of gauge. The natural choice is the Coulomb gauge (radiation gauge),

$$\nabla \cdot \mathbf{A} = 0,$$

since in quantum electro-dynamics this gauge implies that electrons interact via the usual Coulomb potentials and the coupling to the transverse field is minimal, i.e., derivatives are replaced by covariant derivatives.

The interesting quantity is the lowest energy that the system can have.

Definition 1 (Energy).

$$E_{N,K} := \inf \mathcal{E}_0[\Psi, \Psi].$$

where the infimum is taken over all divergence free $\mathbf{A}$ fields, pairwise distinct nuclear locations $\mathbf{R}_1, \dots, \mathbf{R}_K$ and normalized, antisymmetric states $\Psi \in \mathcal{H}_{N,0}$.

4.1. Stability with Small α and Small Particle Number

Since this result is not a main point of this paper we shall be brief—even sketchy. If a single particle Ψ is in the positive spectral subspace of $D(\mathbf{0})$ then the action of $D(\mathbf{0})$ on Ψ is the same as multiplication of each component by $(p^2 + m^2)^{1/2}$ in Fourier space. For such functions we see that $(\Psi, D(\mathbf{0}) \Psi)$ exceeds $(\Psi, |\nabla| \Psi)$, so we may as well replace $D(\mathbf{0})$ by $|\nabla|$ and also drop the condition that Ψ belongs to the positive spectral subspace of $D(\mathbf{0})$.

The next step is to use the lower bound on V_c in (5) so that the energy is now bounded below by a sum of one-body operators, in a manner similar to that in Section 3 with $\tilde{\alpha}_c$ as in Theorem 1:

$$\mathcal{E}_0[\Psi, \Psi] \geq \left(\Psi, \sum_{v=1}^N \left(1 - \frac{\alpha}{\tilde{\alpha}_c} \right) |\nabla_v| \Psi \right) + \sqrt{\alpha} \int_{\mathbb{R}^3} \mathbf{j} \cdot \mathbf{A} + \frac{1}{8\pi} \int_{\mathbb{R}^3} B^2. \quad (16)$$

(Note again that the “effective spin” q is 2, as can be seen by repeating the above argument.) Here $\mathbf{j}(\mathbf{r})$ is the current in the state Ψ and it is trivially bounded above pointwise by the density $\rho(\mathbf{r})$ in the state Ψ (defined in Appendix B). Therefore the integral involving $\mathbf{A}$ is bounded below by

$$\int_{\mathbb{R}^3} \mathbf{j} \cdot \mathbf{A} \geq - \int_{\mathbb{R}^3} \rho A \geq - \|A\|_6 N^{1/3} \|\rho\|_{4/3}^{2/3}.$$

Now $\int B^2 \geq \int |\nabla A|^2$ and this is not less than $K_3^{-2} \|A\|_6^2$ by Sobolev’s inequality where $K_3 = 4^{1/3} (3\pi)^{-1/2} \pi^{-1/6}$ (see ref. 16, p. 367). Similarly, the kinetic energy $(\Psi, \sum_{v=1}^N |\nabla_v| \Psi)$ is bounded below by $1.63q^{-1/3} \int \rho^{4/3}$, which was proved by Daubechies⁽⁶⁾ and which follows from (8). If we use these inequalities and then minimize the energy with respect to the unknown quantity $\|A\|_6$, we easily find that the energy is non-negative as long as

$$1.63(1 - \alpha/\tilde{\alpha}_c) \geq 2\pi N^{2/3} K_3^2 q^{1/3} \alpha$$

with $q = 2$.

We shall show in Subsection 4.2 that the condition that $N^{2/3}\alpha$ is small, which—as we just proved—ensures boundedness from below, is in fact also necessary for the energy to be bounded from below.

4.2. Instability for All α and Large Particle Number

The main result of this section is that there is no stability in this model for any fixed, positive α if N and K are allowed to be arbitrary.

Theorem 2 (Instability). There exists a universal number $C > 0$ such that for all values of the parameters $\alpha > 0$, $m \geq 0$, $K = 1, 2, 3, \dots$, and all values of $N = 1, 2, 3, \dots$, and of $Z_1, Z_2, \dots, Z_K$ satisfying

$$\sum_{\kappa=1}^K Z_{\kappa} > C \max\{\alpha^{-3/2}, 1\}, \quad N > C \max\{\alpha^{-3/2}, 1\}, \quad \sum_{\kappa=1}^K Z_{\kappa}^2 > 2$$

we have

$$E_{N,K} = -\infty.$$

Proof. The theorem follows if, for all $E > 0$, we show the existence of three quantities for which $\mathcal{E}_0[\Psi, \Psi] \leq -E$, with $\Psi = \psi_1 \wedge \cdots \wedge \psi_N$:

- A. A vector potential $\mathbf{A}$ with $\nabla \cdot \mathbf{A} = 0$.
- B. Orthonormal spinors $\psi_1, \dots, \psi_N \in \mathcal{H}_+$.
- C. Nuclear coordinates $\mathbf{R}_1, \dots, \mathbf{R}_K$.

Our construction will depend on four parameters (to be specified at the very end), $\delta > 0$ a momentum scale, which we shall let tend to infinity, $\theta > 0$, which will be chosen sufficiently small (but independently of N), and $P, A_0 > 0$ which will be chosen as functions of N . Finally we denote by $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ the coordinate vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ respectively. We shall use the notation that $\boldsymbol{\omega}_p = \mathbf{p}/|\mathbf{p}|$ is the unit vector in the direction $\mathbf{p} \in \mathbb{R}^3$.

A. The vector potential. We choose the vector potential $\mathbf{A}$ to have Fourier transform

$$\hat{\mathbf{A}}(\mathbf{p}) := A_0 \chi_{B(0, 5\delta)}(\mathbf{p})(\mathbf{n}_2 \cdot \boldsymbol{\omega}_p) \mathbf{n}_3 \times \boldsymbol{\omega}_p,$$

where $\chi_{B(0, 5\delta)}$ denotes the characteristic function (in $\mathbf{p}$ -space) of the ball $B(0, 5\delta)$ centered at 0 with radius 5δ . Note first that $\mathbf{A}$ is real since $\hat{\mathbf{A}}$ is real and $\hat{\mathbf{A}}(\mathbf{p}) = \hat{\mathbf{A}}(-\mathbf{p})$. Moreover, $\mathbf{A}$ is divergence free, i.e., it is in the Coulomb gauge, since $-i \nabla \cdot \hat{\mathbf{A}}(\mathbf{p}) = \mathbf{p} \cdot \hat{\mathbf{A}}(\mathbf{p}) = 0$. We easily estimate the self-energy of the magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ corresponding to $\mathbf{A}$

$$\frac{1}{8\pi} \int (\nabla \times \mathbf{A})^2 = \frac{1}{8\pi} \int_{\mathbb{R}^3} |\mathbf{p} \times \hat{\mathbf{A}}(\mathbf{p})|^2 d\mathbf{p} \leq A_0^2 2^{-1} \int_0^{5\delta} p^4 dp = 2^{-1} 5^4 A_0^2 \delta^5. \quad (17)$$

Finally, we note for later use that

$$\hat{\mathbf{A}}(\mathbf{p}) \cdot \mathbf{n}_1 = -A_0 \chi_{B(0, 5\delta)}(\mathbf{p})(\mathbf{n}_2 \cdot \boldsymbol{\omega}_p)^2. \quad (18)$$

B. The orthonormal spinors. For $\mathbf{p}_0 \in \mathbb{R}^3$ define

$$u_{\mathbf{p}_0}(\mathbf{p}) = \sqrt{3/(4\pi)} \delta^{-3/2} \begin{pmatrix} \chi_{B(0, 5\delta)}(\mathbf{p} - \mathbf{p}_0) \\ 0 \end{pmatrix}. \quad (19)$$

We then have a normalized $\psi_{\mathbf{p}_0} \in \mathcal{H}_+$ given by

$$\widehat{\psi_{\mathbf{p}_0}}(\mathbf{p}) = (2E(p)(E(p) + E(0)))^{-1/2} \begin{pmatrix} (E(p) + E(0)) u_{\mathbf{p}_0}(\mathbf{p}) \\ \mathbf{p} \cdot \boldsymbol{\sigma} u_{\mathbf{p}_0}(\mathbf{p}) \end{pmatrix},$$

where $E(p) = (p^2 + m^2)^{1/2}$. Recall that this is the general form of a spinor in the positive spectral subspace $\mathcal{H}_+$ for the free Dirac operator.

For the sake of simplicity we shall henceforth assume that $m = 0$. We leave it to the interested reader to check the estimates for the general case $m \neq 0$. We shall indeed consider spinors with momenta p such that we have $p^2(m^2)^{-1} \rightarrow \infty$ as $\delta \rightarrow \infty$, i.e., $E(p) \approx p$. It is therefore straightforward to estimate the expressions in the general case $m \neq 0$ by the corresponding expressions for $m = 0$.

In particular, we have, for $m = 0$,

$$\widehat{\psi_{\mathbf{p}_0}}(\mathbf{p}) = 2^{-1/2} \begin{pmatrix} u_{\mathbf{p}_0}(\mathbf{p}) \\ \boldsymbol{\omega}_{\mathbf{p}} \cdot \boldsymbol{\sigma} u_{\mathbf{p}_0}(\mathbf{p}) \end{pmatrix},$$

We shall choose N points $\mathbf{p}_1, \dots, \mathbf{p}_N \in \mathbb{R}^3$ such that the following conditions are satisfied.

1. $\min_{v \neq \mu} |\mathbf{p}_v - \mathbf{p}_\mu| > 2\delta$
2. $P \leq p_v \leq 2P$, for all $v = 1, \dots, N$
3. $\boldsymbol{\omega}_{\mathbf{p}_v} \cdot \mathbf{n}_1 \geq 1 - \theta^2$, for all $v = 1, \dots, N$

Condition 1 ensures that the spinors $\psi_{\mathbf{p}_1}, \dots, \psi_{\mathbf{p}_N}$ are orthonormal. The importance of Conditions 2 and 3 will hopefully become clear below.

In order that Conditions 1, 2 and 3 are consistent with having N points (for large N) we must ensure that N balls of radius δ can be packed into the domain defined by Conditions 2 and 3. Since small enough balls can fill at least half the volume of the given region we simply choose P such that

$$2N \leq \frac{\text{Vol}(\{\mathbf{p} \mid P \leq p \leq 2P, 1 - \theta^2 \leq \boldsymbol{\omega}_{\mathbf{p}} \cdot \mathbf{n}_1\})}{(4\pi/3) \delta^3} = \frac{7}{2} \theta^2 \frac{P^3}{\delta^3}.$$

Note that the assumption that N is larger than some universal number ensures that the balls are small, i.e., that δ is small enough compared to P . Thus, we have the condition

$$P \geq \left(\frac{4N}{7\theta^2} \right)^{1/3} \delta. \quad (20)$$

In particular, since we shall choose θ independently of N we may assume that N is large enough that the above condition implies $P\theta \geq 2\delta$. (Since we shall choose $\delta \rightarrow \infty$ we see that the momenta of the spinors satisfy $p^2 m^{-2} \rightarrow \infty$.)

We are now prepared to calculate $(\Psi, \sum_{v=1}^N D_v(\mathbf{A}) \Psi)$, where $\Psi = \psi_{\mathbf{p}_1} \wedge \cdots \wedge \psi_{\mathbf{p}_N}$. We obtain

$$\left(\Psi, \sum_{v=1}^N D_v(\mathbf{A}) \Psi \right) = \sum_{v=1}^N \left(T_v + \int e \mathbf{J}_v \cdot \mathbf{A} \right), \quad (21)$$

where

$$\mathbf{j}_v(\mathbf{r}) := \psi_{\mathbf{p}_v}^*(\mathbf{r}) \boldsymbol{\alpha} \psi_{\mathbf{p}_v}(\mathbf{r}), \quad (22)$$

is the current of the v -th one-electron state $\psi_{\mathbf{p}_v}$, and

$$T_v := (\psi_{\mathbf{p}_v}, (-i\boldsymbol{\alpha} \cdot \nabla + \beta m) \psi_{\mathbf{p}_v}) = \int E(p) |u_{\mathbf{p}_v}(\mathbf{p})|^2 d\mathbf{p} \leq p_v + \delta, \quad (23)$$

since we have assumed that $m=0$ and hence $E(p) = p$.

We must evaluate the current integral

$$\begin{aligned} \int \mathbf{j}_v \cdot \mathbf{A} &= (2\pi)^{-3/2} 2\Re \iint u_{\mathbf{p}_v}^*(\mathbf{q} - \mathbf{p}) (\hat{\mathbf{A}}(\mathbf{p}) \cdot \boldsymbol{\sigma}) (\boldsymbol{\omega}_{\mathbf{q}} \cdot \boldsymbol{\sigma}) u_{\mathbf{p}_v}(\mathbf{q}) d\mathbf{p} d\mathbf{q} \\ &= (2\pi)^{-3/2} 2\Re \iint [\hat{\mathbf{A}}(\mathbf{q} - \mathbf{p}) \cdot \boldsymbol{\omega}_{\mathbf{q}} u_{\mathbf{p}_v}^*(\mathbf{p}) u_{\mathbf{p}_v}(\mathbf{q}) \\ &\quad + i u_{\mathbf{p}_v}^*(\mathbf{p}) (\hat{\mathbf{A}}(\mathbf{q} - \mathbf{p}) \times \boldsymbol{\omega}_{\mathbf{q}}) \cdot \boldsymbol{\sigma} u_{\mathbf{p}_v}(\mathbf{q})] d\mathbf{p} d\mathbf{q}. \end{aligned}$$

We first observe that

$$\begin{aligned} 2\Re \iint [i u_{\mathbf{p}_v}^*(\mathbf{p}) (\hat{\mathbf{A}}(\mathbf{q} - \mathbf{p}) \times \boldsymbol{\omega}_{\mathbf{q}}) \cdot \boldsymbol{\sigma} u_{\mathbf{p}_v}(\mathbf{q})] d\mathbf{p} d\mathbf{q} \\ = 2\Re \iint [i u_{\mathbf{p}_v}^*(\mathbf{p}) (\hat{\mathbf{A}}(\mathbf{q} - \mathbf{p}) \times \boldsymbol{\omega}_{\mathbf{q}}) \cdot \mathbf{n}_2 \sigma_2 u_{\mathbf{p}_v}(\mathbf{q})] d\mathbf{p} d\mathbf{q} = 0. \end{aligned}$$

The terms containing σ_1 and σ_3 vanish as they are clearly imaginary. The term with σ_2 vanishes because of the choice (19) of $u_{\mathbf{p}_v}$.

Note that $u_{\mathbf{p}_v}^*(\mathbf{p}) u_{\mathbf{p}_v}(\mathbf{q}) = 0$ unless $|\mathbf{p} - \mathbf{q}| < 2\delta$ and $|\mathbf{q} - \mathbf{p}_v| \leq \delta$. Thus $\boldsymbol{\omega}_{\mathbf{p}_v} - \boldsymbol{\omega}_{\mathbf{q}} = p_v^{-1}(\mathbf{p}_v - \mathbf{q}) + \boldsymbol{\omega}_{\mathbf{q}} p_v^{-1}(q - p_v)$ and we obtain for $|\mathbf{p}_v - \mathbf{q}| < 2\delta$ that

$$|\boldsymbol{\omega}_{\mathbf{p}_v} - \boldsymbol{\omega}_{\mathbf{q}}| \leq 2\delta p_v^{-1} \leq 2\delta P^{-1},$$

where we used that $p_v \geq P$. Since $\boldsymbol{\omega}_{\mathbf{p}_v} \cdot \mathbf{n}_1 > 1 - \theta^2$ we have that $|\boldsymbol{\omega}_{\mathbf{p}_v} - \mathbf{n}_1| \leq \theta$ and hence

$$|\hat{\mathbf{A}}(\mathbf{q} - \mathbf{p}) \cdot (\boldsymbol{\omega}_{\mathbf{q}} - \mathbf{n}_1)| \leq (2\delta P^{-1} + \theta) A_0 \leq 2\theta A_0.$$

Hence, since $|\mathbf{p} - \mathbf{q}| < 2\delta$ we get from (18) that

$$\hat{\mathbf{A}}(\mathbf{q} - \mathbf{p}) \cdot \boldsymbol{\omega}_{\mathbf{q}} \leq -A_0[(\boldsymbol{\omega}_{\mathbf{q} - \mathbf{p}} \cdot \mathbf{n}_2)^2 - 2\theta].$$

Thus,

$$\begin{aligned} \int \mathbf{j}_v \cdot \mathbf{A} &= (2\pi)^{-3/2} 2\Re \iint \hat{\mathbf{A}}(\mathbf{q} - \mathbf{p}) \cdot \boldsymbol{\omega}_{\mathbf{q}} u_{\mathbf{p}_v}^*(\mathbf{p}) u_{\mathbf{p}_v}(\mathbf{q}) d\mathbf{q} d\mathbf{p} \\ &\leq \frac{3}{(2\pi)^{5/2}} A_0 \delta^{-3} \iint_{|\mathbf{q}|, |\mathbf{p}| < \delta} [(\boldsymbol{\omega}_{\mathbf{q} - \mathbf{p}} \cdot \mathbf{n}_2)^2 - 2\theta] d\mathbf{p} d\mathbf{q} \\ &= \frac{3}{(2\pi)^{5/2}} A_0 \delta^3 \iint_{|\mathbf{q}|, |\mathbf{p}| < 1} [(\boldsymbol{\omega}_{\mathbf{q} - \mathbf{p}} \cdot \mathbf{n}_2)^2 - 2\theta] d\mathbf{p} d\mathbf{q}. \end{aligned}$$

We now make the choice

$$\theta = \frac{1}{3} \left(\iint_{|\mathbf{q}|, |\mathbf{p}| < 1} 1 d\mathbf{p} d\mathbf{q} \right)^{-1} \iint_{|\mathbf{q}|, |\mathbf{p}| < 1} (\boldsymbol{\omega}_{\mathbf{q} - \mathbf{p}} \cdot \mathbf{n}_2)^2 d\mathbf{p} d\mathbf{q}$$

and arrive at

$$\int \mathbf{j}_v \cdot \mathbf{A} \leq -\frac{4}{3} (2\pi)^{-1/2} A_0 \theta \delta^3.$$

From (21) and (23) we therefore obtain

$$\left(\boldsymbol{\Psi}, \sum_{v=1}^N D_v(\mathbf{A}) \boldsymbol{\Psi} \right) \leq \sum_{v=1}^N [|\mathbf{p}_v| + \delta - \frac{4}{3} (2\pi)^{-1/2} A_0 e \delta^3 \theta]. \quad (25)$$

C. The nuclear coordinates. Finally, we show how to choose the nuclear coordinates following an idea in ref. 18. Consider the electronic density of the state Ψ , $\rho(\mathbf{r}) = \sum_{v=1}^N |\psi_{\mathbf{p}_v}(\mathbf{r})|^2$ then

$$(\Psi, V_c \Psi) \leq - \sum_{k=1}^K \int \frac{Z_k \rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_k|} d\mathbf{r} + D(\rho, \rho) + \sum_{\substack{\kappa, \lambda=1 \\ \kappa < \lambda}}^K \frac{Z_\kappa Z_\lambda}{|\mathbf{R}_\kappa - \mathbf{R}_\lambda|},$$

where we introduced $D(\rho, \rho) := \frac{1}{2} \iint \rho(\mathbf{r}) |\mathbf{r} - \mathbf{r}'|^{-1} \rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}'$.

Note now that $\int N^{-1} \rho = 1$, i.e., $N^{-1} \rho$ can be considered a probability distribution. We may therefore average $(\Psi, V_c \Psi)$ considered as a function of $\mathbf{R}_1, \dots, \mathbf{R}_K$ with respect to the probability measure

$$\mathbf{R}_1, \dots, \mathbf{R}_K \mapsto N^{-1} \rho(\mathbf{R}_1) \cdots N^{-1} \rho(\mathbf{R}_K).$$

We obtain

$$\begin{aligned} & \int (\Psi, V_c \Psi) N^{-1} \rho(\mathbf{R}_1) \cdots N^{-1} \rho(\mathbf{R}_K) d\mathbf{R}_1 \cdots d\mathbf{R}_K \\ &= \left[(1 - (Z/N))^2 - N^{-2} \sum_{\kappa=1}^K Z_\kappa^2 \right] D(\rho, \rho). \end{aligned}$$

We shall prove that $[\dots] < 0$. There are two cases.

(1) $N \geq Z$: By moving electrons to infinity we may assume that $Z \leq N < Z + 1$. Therefore

$$\begin{aligned} \left[(1 - (Z/N))^2 - N^{-2} \sum_{\kappa=1}^K Z_\kappa^2 \right] &\leq Z^{-2} - (Z+1)^{-2} \sum_{\kappa=1}^K Z_\kappa^2 \\ &\leq Z^{-2} - 2(Z+1)^{-2} < 0. \end{aligned}$$

(2) $N < Z$: We may move nuclei to infinity and assume that $N < Z < N + \max_\kappa Z_\kappa$. Therefore

$$\left[(1 - (Z/N))^2 - N^{-2} \sum_{\kappa=1}^K Z_\kappa^2 \right] \leq (\max_\kappa Z_\kappa)^2 N^{-2} - N^{-2} \sum_{\kappa=1}^K Z_\kappa^2 \leq 0.$$

We can therefore find nuclear positions $\mathbf{R}_1, \dots, \mathbf{R}_K$ such that $(\Psi, V_c \Psi) \leq 0$.

Using these coordinates together with (17) and (25) we get

$$\mathcal{E}_0[\Psi, \Psi] \leq \sum_{v=1}^N [p_v + \delta - \frac{4}{3} (2\pi)^{-1/2} A_0 e \delta^3 \theta] + 2^{-1} 5^4 A_0^2 \delta^5.$$

We now choose $P = (4N/7\delta^2)^{1/3} \delta$ in accordance with (20). We then choose A_0 such that $\frac{4}{3}(2\pi)^{-1/2} A_0 e\delta^3\theta = 4P$, i.e.,

$$A_0 = \frac{3}{112^{1/3}} (2\pi)^{1/2} \theta^{-5/3} e^{-1} \delta^{-2} N^{1/3}.$$

Then, using Condition 2 we have $p_v - (8\pi/3) A_0 e\delta^3\theta \leq -2P$. If we insert these values and estimates above we find

$$\mathcal{E}_0[\Psi, \Psi] \leq \delta [-2N^{4/3}(4/(7\theta^2))^{1/3} + N + 3^{25} 4^{112} \pi \theta^{-10/3} \alpha^{-1} N^{2/3}].$$

It is now clear that the expression in [] will be negative for N sufficiently large. The energy can therefore be made arbitrarily negative by choosing δ large.

APPENDIX A. BKS INEQUALITIES

As a convenience to the reader we give a proof of some cases of the inequalities due to Birman, Koplienko, and Solomyak.⁽²⁾ The case needed in Section 3 corresponds to $p=2$ below. There we are interested in $(B-A)_-$, but here we treat $(B-A)_+$ to simplify keeping track of signs. The proof is the same. Recall that $X_+ := (|X| + X)/2$.

Theorem 3. Let $p \geq 1$ and suppose that A and B are two non-negative, self-adjoint linear operators on a separable Hilbert space such that $(B^p - A^p)_+^{1/p}$ is trace class. Then $(B-A)_+$ is also trace class and

$$\text{tr}(B-A)_+ \leq \text{tr}(B^p - A^p)_+^{1/p}.$$

Proof. Our proof will use essentially only two facts: $X \mapsto X^{-1}$ is operator monotone decreasing on the set of nonnegative self-adjoint operators (i.e., $X \geq Y \geq 0 \Rightarrow Y^{-1} \geq X^{-1}$) and $X \mapsto X^r$ is operator monotone increasing on the set of nonnegative self-adjoint operators for all $0 < r \leq 1$. Consequently, $X \mapsto X^{-r}$ is operator monotone decreasing for $0 < r \leq 1$.

As a preliminary remark, we can suppose that $B \geq A$. To see this, write $B^p = A^p + D$. If we replace B by $[A^p + D_+]^{1/p}$ then $(B^p - A^p)_+ = D_+$ is unchanged, while $X := B - A \mapsto [A^p + D_+]^{1/p} - A$ can only get bigger because $X \mapsto X^{1/p}$ is operator monotone on the set of positive operators. Since the trace is also operator monotone, we can therefore suppose that $D = D_+$, i.e., $B^p - A^p + C^p$ with $A, B, C \geq 0$. Our goal is to prove that

$$\text{tr}[(A^p + C^p)^{1/p} - A] \leq \text{tr } C, \tag{26}$$

under the assumption that C is trace class.

To prove (26) we consider the operator $X := [A^p + C^p]^{1/p} - A$, which is well defined on the domain of A . We assume, at first, that $A^p \geq \varepsilon^p$ for some positive number ε . Then, by the functional calculus, and with

$$E := [A^p + C^p]^{(1-p)/p} \quad \text{and} \quad P := A^{1-p} - E$$

we have

$$X = E[A^p + C^p] - A^{1-p}A^p = -PA^p + EC^p \quad (27)$$

Clearly, $P \geq 0$ and $0 \leq P \leq \varepsilon^{1-p}$.

Let $Y := EC^p$. We claim that Y is trace class. This follows from $Y^*Y = C^pE^2C^p \leq C^pC^{2-2p}C^p = C^2$. Thus, $|Y| \leq C$, and hence $\text{tr } Y = \text{tr } C^{p/2}EC^{p/2} \leq \text{tr } C$.

It is also true that P is trace class. To see this, use the integral representation, with suitable $c > 0$, $A^{1-p} = c \int_0^\infty (t + A)^{-1} t^{1-p} dt$. Use this twice and then use the resolvent formula. In this way we find that

$$P = c \int_0^\infty (A^p + t)^{-1} C^p (A^p + C^p + t)^{-1} t^{(1-p)/p} dt.$$

Since C is trace class, so is C^p , and the integral converges because of our assumed lower bound on A . Thus, P is trace class and hence there is a complete, orthonormal family of vectors $v_1, v_2, \dots$, each of which is an eigenvector of P .

Since $X \geq 0$, the trace of X is well defined by $\sum_{j=1}^\infty (v_j, Xv_j)$ for any complete, orthonormal family. The same remark applies to EC^p since it is trace class. Thus, to complete the proof of (26) it suffices to prove that $(v_j, PA^pv_j) \geq 0$ for each j . But this number is $\lambda_j(v_j, A^pv_j) \geq 0$, where λ_j is the (nonnegative) eigenvalue of P , and the positivity follows from the positivity of A .

We now turn to the case of general $A \geq 0$. We can apply the above proof to the operator $A + \varepsilon$ for some positive number ε . Thus we have

$$\text{tr} [(A + \varepsilon)^p + C^p]^{1/p} - (A + \varepsilon) \leq \text{tr } C. \quad (28)$$

Let $\varphi_1, \varphi_2, \dots$ be an orthonormal basis chosen from the domain of A^p . This basis then also belongs to the domain of A and the domain of $[(A + \varepsilon)^p + C^p]^{1/p}$ for all $\varepsilon \geq 0$. We then have

$$\text{tr } X = \sum_j (\varphi_j, X\varphi_j).$$

Note that a-priori we do not know that the trace is finite, but since the operator is non-negative this definition of the trace is meaningful. Operator monotonicity of $X^{1/p}$ gives

$$(\varphi_j, [[(A + \varepsilon)^p + C^p]^{1/p} - (A + \varepsilon)] \varphi_j) \geq (\varphi_j, (X - \varepsilon) \varphi_j).$$

It therefore follows from (28), followed by Fatou's Lemma applied to sums that

$$\begin{aligned} \text{tr } C &\geq \liminf_{\varepsilon \rightarrow 0} \sum_j (\varphi_j, [[(A + \varepsilon)^p + C^p]^{1/p} - (A + \varepsilon)] \varphi_j) \\ &\leq \sum_j \liminf_{\varepsilon \rightarrow 0} (\varphi_j, [[(A + \varepsilon)^p + C^p]^{1/p} - (A + \varepsilon)] \varphi_j) \\ &\geq \text{tr } X. \end{aligned}$$

APPENDIX B. COUNTING SPIN STATES

Our goal here is to prove that when Ψ is in $\mathcal{H}_{N, \mathcal{A}}$, the antisymmetric tensor product of the positive energy subspace of the Dirac operator (with or without a magnetic field, $\mathcal{A}$) then the one-body density matrix is bounded by 2 and not merely by 4, as would be the case if there were no restriction to the positive energy subspace. This result will allow us to use 2 instead of 4 in inequalities (6) and (14). We thank Michael Loss for the idea of this proof.

The one-body density matrix is defined in terms of an N -body density matrix (or function) by the partial trace over $N - 1$ variables. We illustrate this for functions, but the proof works generally. If Ψ is a function, then

$$\Gamma(\mathbf{r}, \sigma; \mathbf{r}', \sigma') := N \int_{(N-1)} \Psi(\mathbf{r}, \sigma, z_2, z_3, \dots, z_N) \overline{\Psi(\mathbf{r}', \sigma', z_2, \dots, z_N)} dz_2 \cdots dz_N,$$

where z denotes a pair $\mathbf{r}, \sigma$ and dz denotes integration over $\mathbb{R}^3$ and summation over the q “spin” states of σ . We are interested in $q = 4$, but that is immaterial for the definition.

The kernel Γ is trace class; in fact its trace is qN . It is also obviously positive definite as an operator. The first remark is that $\Gamma \leq 1$ as an operator. To prove this easily, let ψ be any normalized function of one space-spin variable z and define the function of $N + 1$ variables $\Phi(z_0, \dots, z_N) := \psi(z_0) \Psi(z_1, \dots, z_N) + \sum_{j=1}^N (-1)^j \psi(z_j) \Psi(z_0, \dots, \hat{z}_j, \dots, z_N)$, where $\hat{z}_j$ denotes the absence of z_j . This function Φ is clearly antisymmetric and the integral over all variables of its square is surely nonnegative.

However, this integral is easily computed (using the normalization of ψ and Ψ) to be $(N+1) - (N+1)(\psi, \gamma\psi)$.

The next step is to consider the reduced kernel (without spin) defined by

$$\gamma(\mathbf{r}, \mathbf{r}') := \sum_{\sigma=1}^q \Gamma(\mathbf{r}, \sigma; \mathbf{r}', \sigma),$$

which evidently satisfies the operator inequality $0 \leq \gamma \leq q$, since $\Gamma \leq 1$.

The electron density referred to in Section 4.1 is defined by

$$\rho(\mathbf{r}) := \gamma(\mathbf{r}, \mathbf{r}),$$

but it will not be needed in this Appendix. Another quantity of interest is the current, defined by

$$\mathbf{j}(\mathbf{r}) := \sum_{\sigma, \tau} \Gamma(\mathbf{r}, \sigma; \mathbf{r}, \tau) \mathbf{a}_{\sigma, \tau}.$$

It follows from this that $|\mathbf{j}(\mathbf{r})| \leq \rho(\mathbf{r})$ for every $\mathbf{r} \in \mathbb{R}^3$.

Our goal here is to prove the following fact about γ :

If the N -body Ψ is in $\mathcal{H}_{N, \mathcal{A}}$ then the correspondingly satisfies $0 \leq \gamma \leq 2$ as an operator.

To prove this we introduce the unitary matrix in spin-space (related to the charge conjugation operator)

$$U = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

where 1 denotes the unit 2×2 unit matrix. With a slight abuse of notation, we shall also use U to denote the $U \otimes 1$ acting on the full one-particle space, i.e., $(U)(\mathbf{r}, \sigma') = \sum_{\sigma} U(\sigma', \sigma) f(\mathbf{r}, \sigma)$. The important point to note, which is easily verified from the Dirac equation, is that $\psi \in \mathcal{H}_+$ if and only if $U\psi \in \mathcal{H}_-$, the negative spectral subspace of $D(\mathcal{A})$.

Given $f \in L^2(\mathbb{R}^3)$, we define F^τ to be the spinor $F^\tau(\mathbf{r}, \sigma) := f(\mathbf{r}) \delta_{\sigma, \tau}$. Then evidently $(f, \gamma f) = \sum_{\tau} (F^\tau, \Gamma F^\tau)$. However, since the matrix U merely permutes the spin indices and possibly changes the sign from $+$ to $-$, we have that $\sum_{\tau} (F^\tau, \Gamma F^\tau) = \sum_{\tau} (F^\tau, \Gamma_U F^\tau)$, with $\Gamma_U := U^{-1} \Gamma U$. (Actually, the proof only requires that U be unitary, nothing more.)

We claim that $\Gamma + \Gamma_U \leq 1$ in which case we have proved that $(f, \gamma f) \leq q/2 = 2$, as claimed. To see this, we note that $\Gamma \leq 1$ on $\mathcal{H}_+$ and $\Gamma_U \leq 1$ on $\mathcal{H}_-$. Since the two subspaces are orthogonal, $\Gamma + \Gamma_U \leq 1$ on the whole spinor space.

ACKNOWLEDGMENTS

The authors thank Michael Loss for valuable discussions, especially with regard to Appendix B. After we had proved the results in this paper, including the inequalities in Appendix A, Huzihiro Araki kindly informed us of the paper by Birman, Koplienko, and Solomyak⁽²⁾ in which the inequalities of Appendix A were proved 21 years earlier; we are grateful to him for this help. We are also grateful to M. Griesemer for pointing out several errors in the preliminary version of this paper. The authors also thank the following organizations for their support: the Danish Science Foundation, the European Union, TMR grant FMRX-CT 96-0001, the U.S. National Science Foundation, grant PHY95-13072, and NATO, grant CRG96011.

REFERENCES

1. H. A. Bethe and E. E. Salpeter. Quantum mechanics of one- and two-electron atoms. In S. Flügge, editor, *Handbuch der Physik*, XXXV, pages 88–436. Springer, Berlin, 1 edition, 1957.
2. M. S. Birman, L. S. Koplienko, and M. Z. Solomyak. Estimates for the spectrum of the difference between fractional powers of two self-adjoint operators. *Soviet Mathematics* **19**(3):1–6, 1975. Translation of *Izvestija vyssich*.
3. G. Brown and D. Ravenhall. On the interaction of two electrons. *Proc. Roy. Soc. London A* **208**(A 1095):552–559, September 1951.
4. L. Bugliaro, J. Fröhlich, G. M. Graf, J. Stubbe, and C. Fefferman. A Lieb–Thirring bound for a magnetic Pauli Hamiltonian. *Preprint*, ETH-TH/96-31, 1996.
5. J. G. Conlon. The ground state energy of a classical gas. *Commun. Math. Phys.* **94**(4):439–458, August 1984.
6. I. Daubechies. An uncertainty principle for Fermions with generalized kinetic energy. *Commun. Math. Phys.* **90**:511–520, September 1983.
7. F. J. Dyson and A. Lenard. Stability of matter I. *J. Math. Phys.* **8**:423–434, 1967.
8. F. J. Dyson and A. Lenard. Stability of matter II. *J. Math. Phys.* **9**:698–711, 1967.
9. W. D. Evans, P. Perry, and H. Siedentop. The spectrum of relativistic one-electron atoms according to Bethe and Salpeter. *Commun. Math. Phys.* **18**:733–746, July 1996.
10. P. Federbush. A new approach to the stability of matter problem. I. *J. Math. Phys.* **16**:347–351, 1975.
11. C. Fefferman. Stability of relativistic matter with magnetic fields. *Proc. Nat. Acad. Sci. USA* **92**:5006–5007, 1995.
12. C. Fefferman, J. Fröhlich, and G. M. Graf. Stability of ultraviolet-cutoff quantum electrodynamics with non-relativistic matter. *Texas Math. Phys. Preprint server* 96–379, 1996.
13. J. Fröhlich, E. H. Lieb, and M. Loss. Stability of Coulomb systems with magnetic fields. I: The one-electron atom. *Commun. Math. Phys.* **104**:251–270, 1986.
14. Y. Ishikawa and K. Koc. Relativistic many-body perturbation theory based on the no-pair Dirac–Coulomb–Breit Hamiltonian: Relativistic correlation energies for the noble-gas sequence through Rn ($Z = 86$), the group-IIIB atoms through Hg, and the ions of Ne isoelectronic sequence. *Phys. Rev. A* **50**(6):4733–4742, December 1994.

15. H. J. A. Jensen, K. G. Dyall, T. Saue, and K. Faegri, Jr., Relativistic four-component multiconfigurational self-consistent-field theory for molecules: Formalism. *J. Chem. Physics* **104**(11):4083–4097, March 1996.
16. E. H. Lieb. Sharp constants in the Hardy–Littlewood–Sobolev and related inequalities. *Annals of Mathematics* **118**:349–374, 1983.
17. E. H. Lieb. On characteristic exponents in turbulence. *Commun. Math. Phys.* **92**:473–480, 1984.
18. E. H. Lieb and M. Loss. Stability of Coulomb systems with magnetic fields. II: The many-electron atom and the one-electron molecule. *Commun. Math. Phys.* **104**:271–282, 1986.
19. E. H. Lieb, M. Loss, and H. Siedentop. Stability of relativistic matter via Thomas–Fermi theory. *Helv. Phys. Acta* **69**:974–984, 1996.
20. E. H. Lieb, M. Loss, and J. P. Solovej. Stability of matter in magnetic fields. *Phys. Rev. Lett.* **75**(6):985–989, August 1995.
21. E. H. Lieb and W. E. Thirring. Bound for the kinetic energy of Fermions which proves the stability of matter. *Phys. Rev. Lett.* **35**(11):687–689, September 1975. Erratum: *Phys. Rev. Lett.* **36**(16):11116, October 1975.
22. E. H. Lieb and W. E. Thirring. Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In E. H. Lieb, B. Simon, and A. S. Wightman, editors, *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*. Princeton University Press, Princeton, 1976.
23. E. H. Lieb and H.-T. Yau. The stability and instability of relativistic matter. *Commun. Math. Phys.* **118**:177–213, 1988.
24. J. Sucher. Foundations of the relativistic theory of many-electron atoms. *Phys. Rev. A* **22**(2):348–362, August 1980.
25. J. Sucher. Foundations of the relativistic theory of many-electron bound states. *International Journal of Quantum Chemistry* **25**:3–21, 1984.
26. J. Sucher. Relativistic many-electron Hamiltonians. *Phys. Scripta* **36**:271–281, 1987.
27. W. Thirring, ed. *The Stability of Matter: From Atoms to Stars, Selecta of Elliott H. Lieb*, Springer-Verlag, Berlin, Heidelberg, New York, 1997.
28. E. H. Lieb, H. Siedentop, and J. P. Solovej. Stability of relativistic matter with magnetic fields. *Phys. Rev. Lett.* **79**:1785, 1997.

Part VI

The Thermodynamic Limit
for Real Matter
with Coulomb Forces

The stability of matter

Elliott H. Lieb*

Department of Mathematics and Department of Physics, Princeton University, Princeton, New Jersey 08540

A fundamental paradox of classical physics is why matter, which is held together by Coulomb forces, does not collapse. The resolution is given here in three steps. First, the stability of *atom* is demonstrated, in the framework of nonrelativistic quantum mechanics. Next the Pauli principle, together with some facts about Thomas-Fermi theory, is shown, to account for the stability (i.e., saturation) of *bulk* matter. Thomas-Fermi theory is developed in some detail because, as is also pointed out, it is the asymptotically correct picture of heavy atoms and molecules (in the $Z \rightarrow \infty$ limit). Finally, a rigorous version of screening is introduced to account for *thermodynamic* stability.

CONTENTS

Introduction	553
I. The Stability of Atoms	554
II. Extension of the Uncertainty Principle to Many Fermions	556
III. Thomas-Fermi Theory	557
IV. The Stability of Bulk Matter	562
V. The Thermodynamic Limit	563
References	568

INTRODUCTION

Some features of the physical world are so commonplace that they hardly seem to deserve comment. One of these is that ordinary matter, either in the form of atoms or in bulk, is held together with Coulomb forces and yet is stable. Nowadays we regard this truly remarkable phenomenon as a consequence of quantum mechanics, but it is far from obvious how the conclusion follows from the premise. It is not necessary to ponder the question very long before realizing that it is a subtle one and that the answer is not to be found in any textbook.

Although the Schrödinger equation is half a century old, it was only in the last few years that the proof of stability was completed. The aim of this paper is to present the full story in a simple and coherent way, highlighting only the main physical and mathematical ideas.

The sense of profound unease about the problem just before the dawn of quantum mechanics is exemplified by this quotation (Jeans, 1915):

"... there would be a very real difficulty in supposing that the law $1/r^2$ held down to zero values of r . For the force between two charges at zero distance would be infinite; we should have charges of opposite sign continually rushing together and, when once together, no force would be adequate to separate them... Thus the matter in the universe would tend to shrink into nothing or to diminish indefinitely in size... We should however probably be wrong in regarding a molecule as a cluster of electrons and positive charges. A more likely suggestion, put forward by Larmor and others is that the molecule may consist, in part at least, of rings of electrons in rapid orbital motion."

Jeans' words strike a contemporary chord, especially since one aspect of the problem that worried him has not yet been fully resolved. This is that electrons and nuclei have a magnetic dipole-dipole interaction whose energy goes as r^{-3} . Although the angular average of this interaction vanishes, the interaction can cause the collapse that Jeans feared, even with Schrödinger mechanics. A proper quantum electrodynamics is needed to describe the dipolar interaction at very small distances. For that reason spin dependent forces will be ignored in this paper; only nonrelativistic quantum mechanics will be considered.

It is difficult to find a reliable textbook answer even to the question: How does quantum mechanics prevent the collapse of an atom? One possibility is to say that the Schrödinger equation for the hydrogen atom can be solved and the answer seen explicitly. This is hardly satisfactory for the many-electron atom or for the molecule. Another possible answer is the Heisenberg uncertainty principle. This, unfortunately, is a false argument, as shown in Sec. I. There is, however, a much better uncertainty principle, formulated by Sobolev, which does adequately describe the intuitive fact that a particle's kinetic energy increases sufficiently fast, as the wave function is compressed, to prevent collapse. (See Kato, 1951).

The next question to consider is well stated in this quotation from Ehrenfest (in Dyson, 1967):

"We take a piece of metal. Or a stone. When we think about it, we are astonished that this quantity of matter should occupy so large a volume. Admittedly, the molecules are packed tightly together, and likewise the atoms within each molecule. But why are the atoms themselves so big?... Answer: only the Pauli principle, 'No two electrons in the same state.' That is why atoms are so unnecessarily big, and why metal and stone are so bulky."

Dyson then goes on to say that without the Pauli principle

"We show that not only individual atoms but matter in bulk would collapse into a condensed high-density phase. The assembly of any two macroscopic objects would release energy comparable to that of an atomic bomb."

Two distinct facts are involved here. One is that matter is stable (or saturates), meaning that the ground

*Work partially supported by U. S. National Science Foundation grant MCS 75-21684.

state energy is bounded below by a constant times the first power, and not a higher power, of the particle number. This was proved for the first time by Dyson and Lenard (Dyson and Lenard, 1967, and Lenard and Dyson, 1968), in a beautiful series of papers. Their method is quite complicated, however, and a simpler proof is given in Sec. IV. In addition, they used sufficiently many inequalities that their estimate (for hydrogen atoms) is about -10^{14} Ry/particle. We will obtain a bound of about -23 Ry/particle. The second fact is that matter would definitely not be stable if electrons were bosons (Dyson, 1967). The energy would increase at least as fast as $-N^{7/5}$.

Therefore, Ehrenfest's surmise that the Pauli principle plays a crucial role in preventing collapse is correct. The problem is to display the essence of the Pauli principle in a clear, succinct and mathematically precise way. Unless this is done the physics of stability will remain unclear.

The key fact is developed in Sec. II: If $\rho(x)$ is the one-particle density of any fermion wave function then the total kinetic energy is bounded below by (constant) $\int \rho(x)^{5/3} dx$.¹ This inequality may be termed the uncertainty principle for fermions. It is simple yet powerful enough to establish stability.

Given this bound, it is then necessary to show how the kinetic energy eventually overcomes the r^{-1} Coulomb singularity. It turns out that Thomas-Fermi (TF) theory is exactly what is needed for this purpose because, as Teller discovered in 1962, atoms do not bind in TF theory. Thus TF theory immediately implies saturation. The necessary facts about TF theory are developed in Sec. III.

There is also another good reason for understanding TF theory in detail. The theory used to be regarded as an uncertain approximation in atomic physics, but it is now known that it is more than that. It happens to be an asymptotically correct theory of atoms and molecules as the nuclear charges tend to infinity. In short, TF theory and the theory of the hydrogen atom constitute two opposite, but rigorous foundations for the many electron problem.

After putting together the results of Sec. II and III in Sec. IV, and thereby proving the stability of bulk matter, we address the third main topic of this paper in Sec. V. Does a sensible thermodynamic limit exist for matter? The problem here centers around the long range r^{-1} nature of the Coulomb potential, not the short range singularity. Put another way, the question is that if matter does not implode, how do we know that it does not explode? Normally systems with potentials that fall off less slowly than $r^{-3-\epsilon}$ for some $\epsilon > 0$ cannot be expected to have a thermodynamic limit. The crucial physical fact was discovered by Newton in 1687: outside an isotropic distribution of charge, all the charge appears to be concentrated at the center. This fact is the basis for screening, but to use it a geometric fact about the packing of balls will be needed. Quantum mechanics as such plays almost no role in Sec. V.

¹ $\int f(x) dx$, or simply $\int f$, always denotes a three-dimensional integral.

The content of this paper can be summarized as follows:

- (i) Atoms are stable because of an uncertainty principle,
- (ii) Bulk matter is stable because of a stronger uncertainty principle that holds only for fermions;
- (iii) Thermodynamics exists because of screening.

My hope is that the necessary mathematics, which is presented as briefly as possible, will not obscure these simple physical ideas.

This paper is based on research carried out over the past few years, and it was my good fortune to have had the benefit of collaboration with J. L. Lebowitz, B. Simon, and W. E. Thirring. Without their insights and stimulation probably none of this would have been carried to fruition. Secs. II and IV come from Lieb and Thirring (1975), Sec. III from Lieb and Simon (1977), and Sec. V from Lieb and Lebowitz (1972).

Lectures given in 1976 at the Centro Internazionale Matematico Estivo in Bressanone were the impetus for writing this paper. The bibliography is not intended to be scholarly, but I believe no theorem or idea has been quoted without proper credit.

I am doubly grateful to S. B. Treiman. He kindly invited me to submit this paper to *Reviews of Modern Physics*, and he also generously devoted much time to reading the manuscript and made many valuable suggestions to improve its clarity.

I. THE STABILITY OF ATOMS

By the phrase "stability of an atom" is meant that the ground state energy of an atom is finite. This is a weaker notion than the concept of H stability of matter, to be discussed in Sec. IV, which means that the ground state energy of a many-body system is not merely bounded below but is also bounded by a constant times the number of particles. This, in turn, is different from thermodynamic stability discussed in Sec. V.

Consider the Hamiltonian for the hydrogenic atom:

$$H = -\Delta - Z|x|^{-1} \quad (1)$$

(using units in which $\hbar^2/2 = 1$, $m = 1$, and $|e| = 1$). H acts on $L^2(\mathbb{R}^3)$, the square integrable functions on 3-space. Why is the ground state energy finite, i.e., why is

$$\langle \psi, H\psi \rangle \geq E_0 \langle \psi, \psi \rangle \quad (2)$$

for some $E_0 > -\infty$? The obvious elementary quantum mechanics textbook answer is the *Heisenberg uncertainty principle* (Heisenberg, 1927): If the *kinetic energy* is defined by

$$T_\psi = \int |\nabla \psi(x)|^2 dx, \quad (3)$$

and if

$$\langle |x|^2 \rangle_\psi = \int |x|^2 |\psi(x)|^2 dx,$$

then when

$$\langle \psi, \psi \rangle = \|\psi\|_2^2 = \int |\psi(x)|^2 dx = 1,$$

$$T_\psi \langle |x|^2 \rangle_\psi \geq \frac{3}{4}. \quad (4)$$

The intuition behind applying the Heisenberg uncertainty principle (4) to the ground state problem (2) is that if the electron tries to get within a distance R of the nucleus, the kinetic energy T_ψ is at least as large as R^{-2} . Consequently $\langle \psi, H\psi \rangle \geq R^{-2} - Z/R$, and this has a minimum $-Z^2/4$ for $R = 2/Z$.

The above argument is *false*! The Heisenberg uncertainty principle says no such thing, despite the endless invocation of the argument. Consider a ψ consisting of two parts, $\psi = \psi_1 + \psi_2$. ψ_1 is a narrow wave packet of radius R centered at the origin with $\int |\psi_1|^2 = \frac{1}{2}$. ψ_2 is spherically symmetric and has support in a narrow shell of mean radius L and $\int |\psi_2|^2 = \frac{1}{2}$. If L is large then, roughly, $\int |x|^2 |\psi(x)|^2 dx \sim L^2/2$, whereas $\int |x|^{-1} |\psi(x)|^2 dx \sim 1/2R$. Thus, from (4) we can conclude *only* that $T_\psi > 9/2L^2$ and hence that $\langle \psi, H\psi \rangle \geq 9/2L^2 - Z/2R$. With this wave function, and using *only* the Heisenberg uncertainty principle, we can make E_0 arbitrarily negative by letting $R \rightarrow 0$.

A more colorful way to put the situation is this: an electron cannot have both a sharply defined position and momentum. If one is willing to place the electron in two widely separated packets, however, say here and on the moon, then the Heisenberg uncertainty principle *alone* does not preclude each packet from having a sharp position and momentum.

Thus, while Eq. (4) is correct it is a pale reflection of the power of the operator $-\Delta$ to prevent collapse. A better uncertainty principle (i.e., a lower bound for the kinetic energy in terms of some integral of ψ which does not involve derivatives) is needed, one which reflects more accurately the fact that if one tries to compress a wave function *anywhere* then the kinetic energy will increase. This principle was provided by Sobolev (1938) and for some unknown reason his inequality, which is simple and goes directly to the heart of the matter, has not made its way into the quantum mechanics textbooks where it belongs. Sobolev's inequality in three dimensions [unlike (4) its form is dimension dependent] is

$$T_\psi = \int |\nabla \psi(x)|^2 dx \geq K_s \left\{ \int \rho_\psi(x)^3 dx \right\}^{1/3} = K_s \|\rho_\psi\|_3, \quad (5)$$

where

$$\rho_\psi(x) = |\psi(x)|^2 \quad (6)$$

is the density and

$$K_s = 3(\pi/2)^{4/3} \approx 5.478$$

is known to be the best possible constant. Equation (5) is nonlinear in ρ , but that is unimportant.

A rigorous derivation of (5) would take too long to present but it can be made plausible as follows (Rosen, 1971): K_s is the minimum of

$$K^\psi = \frac{\int |\nabla \psi(x)|^2 dx}{\left\{ \int |\psi(x)|^6 dx \right\}^{1/3}}.$$

Let us accept that a minimizing ψ exists (this is the hard part) and that it satisfies the obvious variational equation

$$-(\Delta \psi)(x) - \alpha \psi(x)^5 = 0$$

with $\alpha > 0$. Assuming also that there is a minimizing ψ

which is nonnegative and spherically symmetric (this can be proved by a rearrangement inequality), one finds by inspection that

$$\psi(x) = (3/\alpha)^{2/3} (1 + |x|^2)^{-1/2}.$$

When this is inserted into the expression for K^ψ the result is $K_s = 3(\pi/2)^{4/3}$. The minimizing ψ is not square integrable, but that is of no concern.

Now let us make a simple calculation to show how good (5) really is. For any ψ

$$\langle \psi, H\psi \rangle \geq K_s \left(\int \rho_\psi(x)^3 dx \right)^{1/3} - Z \int |x|^{-1} \rho_\psi(x) dx = h(\rho), \quad (7)$$

and hence when $\langle \psi, \psi \rangle = 1$

$$\langle \psi, H\psi \rangle \geq \min \left\{ h(\rho): \rho(x) \geq 0, \int \rho = 1 \right\}. \quad (8)$$

The latter calculation is trivial (for any potential) since gradients are not involved. One finds that the solution to the variational equation is $\bar{\rho}(x) = \alpha |x|^{-1} R^{-1} |x|^{1/2}$ for $|x| \leq R$ and $\bar{\rho}(x) = 0$ for $|x| \geq R$, with $R = K_s \pi^{-4/3} Z^{-1}$. Then

$$h(\bar{\rho}) = Z^2 (\pi/2)^{4/3} / K_s = -\frac{4}{3} Z^2 \text{ Ry}.$$

(Recall that one Rydberg $= \text{Ry} = \frac{1}{2}$ in these units.) Thus, Eq. (5) leads easily to the conclusion

$$E_0 \geq -\frac{4}{3} Z^2 \text{ Ry} \quad (9)$$

and this is an excellent lower bound to the correct $E_0 = -Z^2 \text{ Ry}$, especially since no differential equation had to be solved.

In anticipation of later developments, a weaker, but also useful, form of Eq. (5) can be derived. By Hölder's inequality²

$$\int \rho(x)^{5/3} dx \leq \left\{ \int \rho(x)^3 dx \right\}^{1/3} \left\{ \int \rho(x) dx \right\}^{2/3} \quad (10)$$

and, since we always take $\int |\psi|^2 = 1$,

$$T_\psi \geq K_s \int \rho_\psi(x)^{5/3} dx. \quad (11)$$

Note that there is now an exponent 1 outside the integral. Although K_s is the best constant in (5) it is not the best constant in (11). Call the latter K_1 . K_1 is the minimum of

$$\frac{\int |\nabla \psi(x)|^2 dx}{\left\{ \int \rho(x)^{5/3} dx \right\}^{1/3}}$$

subject to $\int \rho(x) dx = 1$. This leads to a nonlinear Schrödinger equation whose numerical solution yields (J. F. Barnes, private communication)

$$K_1 \approx 9.578.$$

In any event

$$K_1 > K^C \approx (\frac{3}{8})(6\pi^2)^{2/3} \approx 9.116,$$

²Hölder's inequality states that

$$\left| \int f(x)g(x) dx \right| \leq \left\{ \int |f(x)|^p dx \right\}^{1/p} \left\{ \int |g(x)|^q dx \right\}^{1/q}$$

when $p^{-1} + q^{-1} = 1$ and $p \geq 1$. To obtain (10) take $f = \rho$, $g = \rho^{2/3}$, $p = 3$, $q = \frac{3}{2}$.

and hence

$$T_\psi \geq K^c \int \rho_\psi(x)^{5/3} dx \text{ when } \int |\psi(x)|^2 dx = 1. \quad (12)$$

K^c is much bigger than K_s ; it is the classical value and will be encountered again in Sec. II and in Sec. III, where its significance will be clarified.

We can repeat the minimization calculation analogous to Eq. (8) using the bound (12) and the functional

$$h^c(\rho) = K^c \int \rho(x)^{5/3} dx - Z \int |x|^{-1} \rho(x) dx.$$

(We could, of course, use the better constant $K_{1,}$) This time

$$\bar{\rho}(x) = \left\{ \frac{3}{5} (Z/K^c) (|x|^{-1} - R^{-1}) \right\}^{3/2} \quad (13)$$

for $|x| \leq R$. R is determined by $\int \rho = 1$ and one finds that $R = (K^c/Z)(4/\pi^2)^{2/3}$ and

$$E_0 \geq -(9Z^2/5K^c)(\pi^2/4)^{2/3} = -3^{1/3}Z^2 \text{ Ry.} \quad (14)$$

The quantity $3^{1/3}$ is only 8.2% greater than $\frac{4}{3}$.

The Sobolev inequality (5) or its variant (12) is, for our purposes, a much better uncertainty principle than Heisenberg's—indeed it is also fairly accurate. We now want to extend (12) to the N -particle case in order to establish the stability of bulk matter. The important new fact that will be involved is that the N particles are *fermions*; that is to say the N -particle wave function is an antisymmetric function of the N -space, spin variables.

II. EXTENSION OF THE UNCERTAINTY PRINCIPLE TO MANY FERMIONS

A well known elementary calculation is that of the lowest kinetic energy, T^V of N fermions in a cubic box of volume V . For large N one finds that

$$T^V \approx q^{-2/3} K^c V \rho^{5/3}, \quad (15)$$

where $\rho = N/V$ and q is the number of spin states available to each particle ($q = 2$ for electrons). Equation (15) is obtained by merely adding up the N/q lowest eigenvalues of $-\Delta$ with Dirichlet ($\psi = 0$) boundary conditions on the walls of the box. The important feature of (15) is that it is proportional to $N^{5/3}$ instead of N , as would be the case if the particles were not fermions. The extra factor $N^{2/3}$ is essential for the stability of matter; if electrons were bosons, matter would not be stable.

Equation (15) suggests that Eq. (12), with a factor $q^{-2/3}$ ought to extend to the N -particle case if $\rho(x)$ is interpreted properly. The idea is old, going back to Lenz (1932), who got it from Thomas-Fermi theory. The proof that something like (12) is not only an approximation but is also a lower bound is new.

To say that the N particles are *fermions* with q spin states means that the N -particle wave function $\psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)$ defined for $x_i \in \mathbb{R}^3$ and $\sigma_i \in \{1, 2, \dots, q\}$ is *antisymmetric* in the pairs (x_i, σ_i) . The norm is given by

$$\langle \psi, \psi \rangle = \sum_{\sigma_i=1}^q \int |\psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 dx_1 \dots dx_N.$$

Define

$$T_\psi = N \sum_{\sigma_i=1}^q \int |\nabla_i \psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 dx_1 \dots dx_N \quad (16)$$

to be the usual kinetic energy of ψ and define

$$\rho_\psi(x) = N \sum_{\sigma_i=1}^q \int |\psi(x, x_2, \dots, x_N; \sigma_1, \dots, \sigma_N)|^2 dx_2 \dots dx_N \quad (17)$$

to be the single particle density, i.e., the probability of finding a particle at x . The analog of (12) is the following (Lieb and Thirring, 1975):

Theorem 1. If $\langle \psi, \psi \rangle = 1$ then

$$T_\psi \geq (4\pi q)^{-2/3} K^c \int \rho_\psi(x)^{5/3} dx. \quad (18)$$

Apart from the annoying factor $(4\pi)^{-2/3} \approx 0.185$, (18) says that the intuition behind considering (15) as a lower bound is correct. We believe that $(4\pi)^{-2/3}$ does not belong in (18) and hope to eliminate it someday. Recent work (Lieb, 1976) has improved the constant by a factor $(1.83)^{2/3} = 1.496$, so we are now off from the conjectured constant $q^{-2/3} K^c$ only by the factor 0.277.

The proof of Theorem 1 is not long but it is slightly tricky. It is necessary first to investigate the negative eigenvalues of a one-particle Schrödinger equation when the potential is nonpositive.

Theorem 2. Let $V(x) \leq 0$ be a potential for the one-particle, three dimensional Schrödinger operator $H = -\Delta + V(x)$ on $L^2(\mathbb{R}^3)$. For $E < 0$ let $N_E(V)$ be the number of eigenstates of H with energies $\leq E$. Then

$$N_E(V) \leq (4\pi)^{-1} (2|E|)^{-1/2} \int |V(x) - E/2|^2 dx, \quad (19)$$

where $|f(x)|_- = |f(x)|$ if $f(x) \leq 0$ and $|f(x)|_- = 0$ otherwise.

Corollary. If $e_1 \leq e_2 \leq \dots \leq 0$ are the negative eigenvalues of H (if any) then

$$\sum_{j=1}^{\infty} |e_j| \leq \frac{4\pi}{15\pi^2} \int |V(x)|^2 dx. \quad (20)$$

Proof. $\sum |e_j| = \int_0^\infty N_{-\alpha}(V) d\alpha$. Insert (19) and do the α integration first and then the x integration. The result is (20). ■

We believe the factor (4π) does not belong in (20).

Proof of Theorem 2. From the Schrödinger equation $H\psi = e\psi$ it is easy to deduce that $N_E(V)$ is equal to the number of eigenvalues which are ≥ 1 of the positive definite Birman-Schwinger operator (Birman, 1961; Schwinger, 1961)

$$B_E(V) = |V|^{1/2} (-\Delta - E)^{-1} |V|^{1/2}. \quad (21)$$

Essentially Eq. (21) comes from the fact that if $H\psi = e\psi$ then $(-\Delta - e)\psi = |V|\psi$. If one defines $|V|^{1/2}\psi \equiv \phi$, then $B_E\phi = \phi$. Thus B_E has an eigenvalue 1 when e is an eigenvalue. However, B_E is a compact positive semidefinite operator on $L^2(\mathbb{R}^3)$ for $E < 0$ and, as an operator, B_E is monotone increasing in E . Thus, if B_E has k eigenvalues ≥ 1 , there exist k numbers $e_1 \leq e_2 \leq \dots \leq e_k \leq E$ such that

B_E has eigenvalue 1.

Consequently $N_E(V) \leq \text{Tr} B_E(V)^2$. On the other hand, $N_E(V) \leq N_{E/2}(-|V - E/2|_-)$ by the variational principle (draw a graph of $V(x) - E/2$). Thus, since $B_E(V)$ has a kernel, $B_E(x, y) = |V(x)|^{1/2} \exp[-|E|^{1/2}|x-y|] [4\pi|x-y|]^{-1} \times |V(y)|^{1/2}$, one has that

$$\begin{aligned} N_E(V) &\leq \text{Tr} B_{E/2}(-|V - E/2|_-)^2 \\ &= (4\pi)^{-2} \int \int dx dy |V(x) - E/2|_- |V(y) - E/2|_- \\ &\quad \times \exp[-(2|E|)^{1/2}|x-y|] |x-y|^{-2}. \end{aligned} \quad (22)$$

Equation (19) results from applying Young's inequality³ to Eq. (22). Alternatively, one can do the convolution integral by Fourier transforms and note that the Fourier transform of the last factor has a maximum at $p=0$, where it is $4\pi(2|E|)^{-1/2}$. ■

Using (20), which is a statement about the energy levels of a single particle Hamiltonian, we can, surprisingly, prove Theorem 1, which refers to the kinetic energy of N fermions.

Proof of Theorem 1. ψ and hence $\rho_\psi(x)$ are given. Consider the non-positive single particle potential $V(x) \equiv -\alpha\rho_\psi(x)^{2/3}$ where α is given by $(2/3\pi)q\alpha^{3/2} = 1$. Next consider the following N -particle Hamiltonian:

$$\tilde{H}_N = \sum_{i=1}^N h_i; \quad h_i = -\Delta_i + V(x_i)$$

on $L^2(\mathbb{R}^3; \mathbb{C})^N$. If E_0 is the fermion ground state energy of $\tilde{H}_N$, we have that $E_0 \geq q \sum e_j$, where the e_j are the negative eigenvalues of the single particle Hamiltonian h . (We merely fill the lowest negative energy levels q times until the N particles are accounted for; if there are k such levels and if $N < kq$ then $E_0 > q \sum e_j$. If $N > kq$, the surplus particles can be placed in wave packets far away from the origin with arbitrarily small kinetic energy.) On the other hand, $E_0 \leq \langle \psi, \tilde{H}_N \psi \rangle = T_\psi - \alpha \int \rho_\psi(x)^{5/3} dx$ by the variational principle. If these two inequalities are combined together with (20), which says that $\sum e_j \geq -(4/15\pi)\alpha^{5/2} \int \rho_\psi(x)^{5/3} dx$, then (18) is the result. ■

It might not be too much out of place to explain at this point why K^c is called the classical constant. The name does not stem from its antiquity, as in the ideal gas kinetic energy (15), but rather from classical mechanics—more precisely the semiclassical approximation to quantum mechanics. This intuitive idea is valuable.

As the proof of Theorem 1 shows, the constant in (18) for T_ψ is simply related to the constant in (20) for the

sum of the eigenvalues. The point is that the semiclassical approximation to this sum is

$$\sum |e_j| \approx (15\pi^2)^{-1} \int |V(x)|^{5/2} dx,$$

and this, in turn, would yield (18) without the $(4\pi)^{-2/3}$ factor. The semiclassical approximation is obtained by saying that a region of volume $(2\pi)^3$ in the six-dimensional phase space (p, x) can accommodate one eigenstate. Hence, integrating over the set $\theta(H)$, in which $H(p, x) = p^2 + V(x)$ is negative,

$$\begin{aligned} \sum e_j &\approx (2\pi)^{-3} \int \int_{\theta(H)} dx dp \{p^2 + V(x)\} \\ &= (2\pi)^{-3} \int dx 4\pi \int_0^{|V(x)|^{1/2}} p^2 dp \{p^2 + V(x)\} \\ &= -(15\pi^2)^{-1} \int |V(x)|^{5/2} dx. \end{aligned}$$

If a coupling constant g is introduced, and if V is replaced by gV , then it is a theorem that the semiclassical approximation is asymptotically exact as $g \rightarrow \infty$ for any V in $L^{5/2}(\mathbb{R}^3)$.

Theorem 1 gives a lower bound to the kinetic energy of fermions which is crucial for the H stability of matter as developed in Sec. IV. To appreciate the significance of Theorem 1 it should be compared with the one-particle Sobolev bound (12). Suppose that $\rho(x) = 0$ outside some fixed domain Ω of volume V . Then since

$$\int_\Omega \rho(x)^{5/3} dx \geq \left\{ \int_\Omega \rho(x) dx \right\}^{5/3} \left\{ \int_\Omega 1 \right\}^{-2/3} = N^{5/3} V^{-2/3}$$

by Hölder's inequality, one sees that T_ψ grows at least as fast as $N^{5/3}$. Using Eq. (12) alone, one would only be able to conclude that T_ψ grows as N . This distinction stems from the Pauli principle, i.e., the antisymmetric nature of the N -particle wave function. As we shall see, this $N^{5/3}$ growth is essential for the stability of matter because without it the ground state energy of N particles with Coulomb forces would grow at least as fast as $-N^{7/5}$ instead of $-N$.

The Fermi pressure is needed to prevent a collapse, but to learn how to exploit it we must first turn to another chapter in the theory of Coulomb systems, namely Thomas-Fermi theory.

III. THOMAS-FERMI THEORY

The statistical theory of atoms and molecules was invented independently by Thomas and Fermi (Thomas, 1927; Fermi, 1927). For many years the TF theory was regarded as an uncertain approximation to the N -particle Schrödinger equation and much effort was devoted to trying to determine its validity (e.g., Gombás, 1949). It was eventually noticed numerically (Sheldon, 1955) that molecules did not appear to bind in this theory, and then Teller (1962) proved this to be a general theorem.

It is now understood that TF theory is really a large Z theory (Lieb and Simon, 1977); to be precise it is exact in the limit $Z \rightarrow \infty$. For finite Z , TF theory is qualita-

³Young's inequality states that

$$\begin{aligned} \left| \iint f(x)g(x-y)h(y) dx dy \right| &\leq \left\{ \int |f(x)|^p dx \right\}^{1/p} \\ &\quad \times \left\{ \int |g(x)|^q dx \right\}^{1/q} \\ &\quad \times \left\{ \int |h(x)|^r dx \right\}^{1/r} \end{aligned}$$

when $p^{-1} + q^{-1} + r^{-1} = 2$ and $p, q, r \geq 1$. For (22) take $p=r=2$ and $q=1$.

tively correct in that it adequately describes the bulk of an atom or molecule. It is not precise enough to give binding. Indeed, it should not do so because binding in TF theory would imply that the cores of atoms bind, and this does not happen. Atomic binding is a fine quantum effect. Nevertheless, TF theory deserves to be well understood because it is exact in a limit; the TF theory is to the many-electron system as the hydrogen atom is to the few-electron system. For this reason the main features of the theory are presented here, mostly without proof.

A second reason for our interest in TF theory is this: in the next section the problem of the H stability of matter will be reduced to a TF problem. The knowledge that TF theory is H stable (this is a corollary of the no binding theorem) will enable us to conclude that the true quantum system is H stable.

The Hamiltonian for N electrons with k static nuclei of charges $z_i > 0$ and locations R_i is

$$H_N = \sum_{i=1}^N -\Delta_i - V(x_i) + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} + U(\{z_j, R_j\}_{j=1}^k), \quad (23)$$

where

$$V(x) = \sum_{j=1}^k z_j |x - R_j|^{-1}, \quad (24a)$$

and

$$U(\{z_j, R_j\}_{j=1}^k) = \sum_{1 \leq i < j \leq k} z_i z_j |R_i - R_j|^{-1}. \quad (24b)$$

The nuclear-nuclear repulsion U is, of course, a constant term in H_N but it is included for two reasons:

(i) We wish to consider the dependence on the R_i of

$$E_N^Q(\{z_j, R_j\}_{j=1}^k) \equiv \text{the ground state energy of } H_N. \quad (25)$$

(ii) Without U the energy will not be bounded by N .

The nuclear kinetic energy is not included in H_N . For the H -stability problem we are only interested in finding a lower bound to E_N^Q , and the nuclear kinetic energy adds a positive term. In other words,

$$\inf_{\{R_i\}} E_N^Q(\{z_j, R_j\}_{j=1}^k)$$

is smaller than the ground state energy of the true Hamiltonian [defined in Eq. (58)] in which the nuclear kinetic energy is included. Later on when we do the proper thermodynamics of the whole system we shall have to include the nuclear kinetic energy.

The problem of estimating E_N^Q is as old as the Schrödinger equation. The TF theory, as interpreted by Lenz (1932), reads as follows: For fermions having q spin states ($q=2$ for electrons) define the TF energy func-

$$\mathcal{E}(\rho) = q^{-2/3} K^c \int \rho(x)^{5/3} - \int V(x) \rho(x) + \frac{1}{2} \iint \rho(x) \rho(y) |x - y|^{-1} dx dy + U(\{z_j, R_j\}_{j=1}^k) \quad (26)$$

for non-negative functions $\rho(x)$. Then for $\lambda \geq 0$

$$E_\lambda^{\text{TF}} \equiv \inf \left\{ \mathcal{E}(\rho) : \int \rho(x) dx = \lambda \right\} \quad (27)$$

is the TF energy for λ electrons (λ need not be an integer, of course). When $\lambda = N$, the minimizing ρ is supposed to approximate the ρ_ϕ given by (17), wherein ψ is the true ground state wave function, and E_N^{TF} is supposed to approximate E_N^Q .

The intuitive idea behind TF theory is this: If ψ is any fermion wave function and T_ψ and ρ_ψ are given by Eqs. (16) and (17), then the first term in (26) is supposed to approximate T_ψ . This is based on the box kinetic energy (15). The last three terms in (26) represent, respectively, the electron-nuclear, electron-electron, and nuclear-nuclear Coulomb energy. E_λ^{TF} in (27) is then the "ground state energy" of (26).

The second and fourth terms on the right side of (26) are exact but the first and third are not. The first is to some extent justified by the kinetic energy inequality, Theorem 1; the third term will be discussed later. In any event, Eqs. (26) and (27) define TF theory.

It would be too much to try to reproduce here the details of our analysis of TF theory. A short summary of some of the main theorems will have to suffice.

The first question is whether or not E_λ^{TF} (which, by simple estimates using Young's and Hölder's inequalities, can be shown to be finite for all λ) is a minimum as distinct from merely an infimum. The distinction is crucial because the TF equation [the Euler-Lagrange equation for (26) and (27)]

$$\frac{5}{3} K^c q^{-2/3} \rho^{2/3}(x) = \max\{\phi(x) - \mu, 0\} \quad (28)$$

with

$$\phi(x) = V(x) - \int \rho(y) |x - y|^{-1} dy \quad (29)$$

has a solution with $\int \rho = \lambda$ if and only if there is a minimizing ρ for E_λ^{TF} . The basic theorem is as follows.

Theorem 3. If $\lambda \leq Z = \sum_{j=1}^k z_j$ then

(i) $\mathcal{E}(\rho)$ has a minimum on the set $\int \rho(x) dx = \lambda$.

(ii) This minimizing ρ (call it ρ_λ^{TF}) is unique and satisfies (28) and (29). μ is non-negative, and $-\mu$ is the chemical potential, i.e.,

$$-\mu = dE_\lambda^{\text{TF}}/d\lambda. \quad (30)$$

(iii) There is no other solution to (28) and (29) (for any μ) with $\int \rho = \lambda$ other than ρ_λ^{TF} .

(iv) When $\lambda = Z$, $\mu = 0$. Otherwise $\mu > 0$, i.e., E_λ^{TF} is strictly decreasing in λ .

(v) As λ varies from 0 to Z , μ varies continuously from $+\infty$ to 0.

(vi) μ is a convex, decreasing function of λ .

(vii) $\phi_\lambda^{\text{TF}}(x) > 0$ for all x and λ . Hence when $\lambda = Z$

$$\frac{5}{3} K^c q^{-2/3} \rho_\lambda^{\text{TF}}(x)^{2/3} = \phi_\lambda^{\text{TF}}(x).$$

If $\lambda > Z$ then $E_\lambda^{\text{TF}}(\lambda)$ is not a minimum and (28) and (29) have no solution with $\int \rho = \lambda$. Negative ions do not exist in TF theory. Nevertheless, E_λ^{TF} exists and $E_\lambda^{\text{TF}} = E_Z^{\text{TF}}$ for $\lambda \geq Z$.

The proof of Theorem 3 is an exercise in functional

analysis. Basically, one first shows that $\mathcal{E}(\rho)$ is bounded below so that E_λ^{TF} exists. The Banach–Alaoglu theorem is used to find an $L^{5/3}$ weakly convergent sequence of ρ 's such that $\mathcal{E}(\rho)$ converges to E_λ^{TF} . Then one notes that $\mathcal{E}(\rho)$ is weakly lower semicontinuous so that a minimizing ρ exists under the subsidiary condition that $\int \rho \leq \lambda$. The uniqueness comes from an important property of $\mathcal{E}(\rho)$, namely that it is *convex*. This also implies that the minimizing ρ satisfies $\int \rho = \lambda$.

A major point to notice is that a solution of the TF equation is obtained as a by-product of minimizing $\mathcal{E}(\rho)$; a direct proof that the TF equation has a solution would be very complicated.

Only in the case $\lambda \leq Z$ is $\rho_\lambda^{\text{TF}}(x)$ positive for all x , when $\lambda < Z$, $\mu > 0$ and, since $\phi_\lambda^{\text{TF}}(x)$ goes to zero as $|x|$ goes to infinity, Eq. (28) implies that $\rho_\lambda^{\text{TF}}(x)$ vanishes outside some bounded set.

Apart from the details presented in Theorem 3, the main point is that TF theory is well defined. In particular the density ρ^{TF} is unique—a state of affairs in marked contrast to that of Hartree–Fock theory (Hartree, 1927–28; Fock, 1930; Slater, 1930; Lieb and Simon, 1973).

The TF density ρ_λ^{TF} has the following properties:

Theorem 4. If $\lambda \leq Z$ then

$$(i) \left(\frac{5}{3}\right) K^\epsilon q^{-2/3} \rho_\lambda^{\text{TF}}(x)^{2/3} \sim z_i |x - R_i|^{-1} \quad (31)$$

near each R_i .

$$(ii) \text{ In the neutral case, } \lambda = Z = \sum_{j=1}^k z_j, \quad (32)$$

$$|x|^3 \rho_\lambda^{\text{TF}}(x) \rightarrow (3/\pi)^{1/3} \left[\left(\frac{5}{3}\right) K^\epsilon q^{-2/3} \right]^{3/2}$$

as $|x| \rightarrow \infty$, irrespective of the distribution of the nuclei.

(iii) $\phi_\lambda^{\text{TF}}(x)$ and $\rho_\lambda^{\text{TF}}(x)$ are real analytic in x away from all the R_i , on all of 3-space in the neutral case and on $\{x: \phi_\lambda^{\text{TF}}(x) > \mu\}$ in the positive ionic case.

Equation (32) is especially remarkable: at large distances one loses all knowledge of the nuclear charges and configuration. Property (i) recalls the singularity found in the minimization of $h^c(\rho)$ [see Eq. (13)].

Equation (31) can be seen from (28) and (29) by inspection. Equation (32) is more subtle but it is consistent with the observation that (28) and (29) can be rewritten (when $\mu = 0$) as

$$-(4\pi)^{-1} \Delta \phi_\lambda^{\text{TF}}(x) = -\left\{ \left(\frac{5}{3}\right) q^{2/3} \phi_\lambda^{\text{TF}}(x) / K^\epsilon \right\}^{3/2}$$

away from the R_i . If it is *assumed* that $\phi_\lambda^{\text{TF}}(x)$ goes to zero as a power of $|x|$ then (32) follows. This observation was first made by Sommerfeld (1932). The proof that a power law falloff actually occurs is somewhat subtle and involves potential theoretic ideas such as that used in the proof of Lemma 8.

As pointed out earlier, the connection between TF theory and the Schrödinger equation is best seen in the limit $Z \rightarrow \infty$. Let the number, k , of nuclei be held fixed, but let $N \rightarrow \infty$ and $z_i \rightarrow \infty$ in such a way that the degree of ionization N/Z is constant, where

$$Z = \sum_{j=1}^k z_j.$$

To this end we make the following definition: Fix $\{z_j, R_j\}_{j=1}^k$ and λ . It is not necessary to assume that λ

$\leq Z$. For each $N = 1, 2, \dots$ define a_N by $\lambda a_N = N$. In H_N (23) replace z_j by $z_j a_N$ and R_j by $R_j a_N^{-1/3}$. This means that the nuclei come together as $N \rightarrow \infty$. If they stay at fixed positions then that is equivalent, in the limit, to isolated atoms, i.e., it is equivalent to starting with all the nuclei infinitely far from each other. Finally, for the nuclear configuration $\{a_N z_j, a_N^{-1/3} R_j\}_{j=1}^k$ let ψ_N be the ground state wave function, E_N^0 the ground state energy, and $\rho_N^0(x)$ be the single particle density as defined by Eq. (17).⁴

It is important to note that there is a simple and obvious scaling relation for TF theory, namely

$$E_{\lambda a}^{\text{TF}}(\{a z_j, a^{-1/3} R_j\}_{j=1}^k) = a^{7/3} E_\lambda^{\text{TF}}(\{z_j, R_j\}_{j=1}^k) \quad (33)$$

and the densities for the two systems are related by

$$\rho_{\lambda a}^{\text{TF}}(a^{-1/3} x) = a^2 \rho_\lambda^{\text{TF}}(x) \quad (34)$$

for any $a \geq 0$. Hence, for the above sequence of systems parametrized by a_N ,

$$a_N^{-7/3} E_N^{\text{TF}}(\{a_N z_j, a_N^{-1/3} R_j\}_{j=1}^k) = E_\lambda^{\text{TF}}(\{z_j, R_j\}_{j=1}^k), \quad (35)$$

$$a_N^2 \rho_N^{\text{TF}}(a_N^{-1/3} x) = \rho_\lambda^{\text{TF}}(x) \quad (36)$$

for all N .

If, on the other hand, the nuclei are held fixed then one can prove that

$$\lim_{N \rightarrow \infty} a_N^{-7/3} E_N^{\text{TF}}(\{a_N z_j, R_j\}) = \sum_{j=1}^k E_{\lambda_j}^{\text{TF}}(z_j), \quad (37)$$

where $E_{\lambda_j}^{\text{TF}}(z)$ is the energy of an isolated atom of nuclear charge z . The λ_j are determined by the condition that $\sum_{j=1}^k \lambda_j = \lambda$ if $\lambda \leq Z$ (otherwise, $\sum_{j=1}^k \lambda_j = Z$) and that the chemical potentials of the k atoms are all the same. Another way to say this is that the λ_j minimize the right side of Eq. (37). With the nuclei *fixed*, the analog of (36) is

$$\lim_{N \rightarrow \infty} a_N^2 \rho_N^{\text{TF}}(a^{-1/3}(x - R_j)) = \rho_{\lambda_j}^{\text{TF}}(x). \quad (38)$$

The right side of Eq. (38) is the ρ for a single atom of nuclear charge z and electron charge λ_j . Equations (37) and (38) are a precise statement of the fact that isolated atoms result from fixing the R_j .

The TF energy for an isolated, *neutral atom* of nuclear charge Z is found numerically to be

$$E_Z^{\text{TF}} = -(2.21) q^{2/3} (K^\epsilon)^{-1} Z^{7/3}. \quad (39)$$

For future use, note that E_Z^{TF} is proportional to $1/K^\epsilon$. Thus, if one considers a TF theory with K^ϵ replaced by some other constant $\alpha > 0$, as will be necessary in Sec. IV, then Eq. (39) is correct if K^ϵ is replaced by α .

Theorem 5. With $a_N = N/\lambda$ and $\{z_j, R_j\}_{j=1}^k$ fixed

- (i) $a_N^{-7/3} E_N^0(\{a_N z_j, a_N^{-1/3} R_j\}_{j=1}^k)$ has a limit as $N \rightarrow \infty$.
- (ii) This limit is $E_\lambda^{\text{TF}}(\{z_j, R_j\}_{j=1}^k)$.
- (iii) $a_N^{-7/3} E_N^0(\{a_N z_j, R_j\}_{j=1}^k)$ has a limit as $N \rightarrow \infty$. This limit is the right side of (37).
- (iv) $a_N^2 \rho_N^0(a_N^{-1/3} x; \{a_N z_j, a_N^{-1/3} R_j\}_{j=1}^k)$ also has a limit as

⁴If E_N^0 is degenerate, ψ_N can be *any* ground state wave function as far as Theorem 5 is concerned. If E_N^0 is not an eigenvalue, but merely inf spec H_N , then it is possible to define an approximating sequence ψ_N , with ρ_N^0 still given by Eq. (17), in such a way that Theorem 5 holds. We omit the details of this construction here.

$N \rightarrow \infty$. If $\lambda \leq Z = \sum_{j=1}^N z_j$, this limit is $\rho_N^{\text{TF}}(x)$ and the convergence is in weak $L^1(\mathbb{R}^3)$. If $\lambda > Z$, the limit is $\rho_N^{\text{TF}}(x)$ in weak $L^1_{\text{loc}}(\mathbb{R}^3)$.

(v) For fixed nuclei, $a_N^2 \rho_N^Q(a_N^{1/3}(x - R_j); \{a_N z_j, R_j\}_{j=1}^N)$ has a limit [in the same sense as (iv)] which is the right side of (38).

The proof of Theorem 5 does not use anything introduced so far. It is complicated, but elementary. One partitions 3-space into boxes with sides of order $Z^{-1/3}$. In each box the potential is replaced by its maximum (respectively, minimum) and one obtains an upper (respectively, lower) bound to E_N^Q by imposing Dirichlet ($\psi = 0$) (respectively, Neumann ($\nabla\psi = 0$)) boundary conditions on the boxes. The upper bound is essentially a Hartree-Fock calculation. The $-r^{-1}$ singularity near the nuclei poses a problem for the lower bound, and it is tamed by exploiting the concept of angular momentum barrier.

What Theorem 5 says, first of all, is that the true quantum energy has a limit on the order of $Z^{7/3}$ when the ratio of electron to nuclear charge is held fixed. Second, this limit is given correctly by TF theory as is shown in Eq. (35). The requirement that the nuclei move together as $Z^{-1/3}$ should be regarded as a refinement rather than as a drawback, for if the nuclei are fixed a limit also exists but it is an uninteresting one of isolated atoms.

Theorem 5 also says that the density ρ_N^Q is proportional to Z^2 and has a scale length proportional to $Z^{-1/3}$. If $\lambda > Z$, Theorem 5 states that the surplus charge moves off to infinity and the result is a neutral molecule. This means that large atoms or molecules cannot have a negative ionization proportional to the total nuclear charge; at best they can have a negative ionization which is a vanishingly small fraction of the total charge. This result is physically obvious for electrostatic reasons, but it is nice to have a proof of it.

Theorem 5 also resolves certain "anomalies" of TF theory:

(a) In real atoms or molecules the electron density falls off exponentially, while in TF theory (Theorem 4) the density falls off as $|x|^{-6}$.

(b) The TF atom shrinks in size as $Z^{-1/3}$ [cf. Eq. (36)] while real large atoms have roughly constant size.

(c) In TF theory there is no molecular binding, as we shall show next, but there is binding for real molecules.

(d) In real molecules the electron density is finite at the nuclei, but in TF theory it goes to infinity as $z_j |x - R_j|^{-3/2}$ (Theorem 4).

As Theorem 5 shows, TF theory is really a theory of heavy atoms or molecules. A large atom looks like a stellar galaxy, poetically speaking. It has a core which shrinks as $Z^{-1/3}$ and which contains most of the electrons. The density (on a scale of Z^2) is not finite at the nucleus because, as the simplest Bohr theory shows, the S -wave electrons have a density proportional to Z^3 which is infinite on a scale of Z^2 . Outside the core is a mantle in which the density is proportional to (cf. Theorem 4) $(3/\pi)^3 (\frac{2}{3})^3 K^2 2^{2/3} Z^2 / (Z^{1/3} |x|)^6$, which is independent of Z ! This density is correct to infinite distances on a length scale $Z^{-1/3}$. The core and the man-

tle contain 100% of the electrons as $Z \rightarrow \infty$. The third region is a transition region to the outer shell, and while it may contain many electrons, it contains only a vanishingly small fraction of them. The fourth region is the outer shell in which chemistry and binding takes place. TF theory has nothing to say about this region. The fifth region is the one in which the density drops off exponentially.

Thus, TF theory deals only with the core and the mantle in which the bulk of the energy and the electrons reside. There ought not to be binding in TF theory, and indeed there is none, because TF energies are proportional to $Z^{7/3}$ and binding energies are of order one. The binding occurs in the fourth layer.

An important question is what is the next term in the energy beyond the $Z^{7/3}$ term of TF theory. Several corrections have been proposed: (e.g., Dirac, 1930; Von Weizsäcker, 1935; Kirzhnits, 1957; Kompaneets and Pavlovskii, 1956; Scott, 1952). With the exception of the last, all these corrections are of order $Z^{5/3}$. Scott (as late as 1952!) said there should be a $Z^{6/3}$ correction because TF theory is not able to treat correctly the innermost core electrons. Let us give a heuristic argument. Recall that in Bohr theory each inner electron alone has an energy proportional to Z^2 . As these inner electrons are unscreened, their energies should be independent of the presence or absence of the electron-electron repulsion. In other words, the Z^2 correction for a molecule should be precisely a sum of corrections, one for each atom. The atomic correction should be the difference between the Bohr energy and the $Z^{7/3}$ TF energy for an atom in which the electron-electron repulsion is neglected. We already calculated the TF energy for such an "atom" in Eq. (14) (put $Z=1$ there and then use scaling; also replace K^c by $q^{-2/3} K^c$). Thus, for a neutral atom without electron-electron repulsion

$$\tilde{E}_Z^{\text{TF}} = -(3^{1/3}/4) q^{2/3} Z^{7/3}. \quad (40)$$

For the Bohr atom, each shell of energy $-Z^2/4n^2$ has n^2 states, so

$$\frac{Z}{q} = \sum_{n=1}^L n^2 + (L+1)^2 \phi = \frac{L^3}{3} + \frac{L^2}{2} + \frac{L}{6} + (L+1)^2 \phi$$

with $0 \leq \phi \leq 1$ being the fraction of the $(L+1)$ th shell that is filled. One finds $L \approx (3Z/q)^{1/3} - \frac{1}{2} - \phi + o(1)$ and

$$E_Z^{\text{Bohr}} = -\frac{Z^2}{4} q \left\{ \phi + \sum_{n=1}^L 1 \right\} \approx \tilde{E}_Z^{\text{TF}} + \frac{q}{8} Z^2.$$

Thus, to the next order, the energy should be

$$E_N^Q(\{z_j, R_j\}_{j=1}^N) = E_N^{\text{TF}}(\{z_j, R_j\}_{j=1}^N) + \frac{1}{8} \sum_{j=1}^N z_j^2 + \text{lower order}, \quad (41)$$

since $q=2$ for electrons. Note that $E_N^{\text{TF}} \sim q^{2/3}$ while the Scott correction is proportional to q .

It is remarkable that Eq. (41) gives a *precise conjecture* about the next correction. It is simple to understand physically, yet we do not have the means to prove it.

The third main fact about TF theory is that there is no binding. This was proved by Teller in 1962. Considering the effort that went into the study of TF theory since its inception in 1927, it is remarkable that the no

binding phenomenon was not seriously noticed until the computer study of Sheldon in 1955. Teller's original proof involved some questionable manipulation with δ functions and for that reason his result was questioned. His ideas were basically right, however, and we have made them rigorous.

Theorem 6 (no binding). If there are at least two nuclei, write the nuclear attraction $V(x) = \sum_{j=1}^k z_j |x - R_j|^{-1}$ as the sum of two pieces, $V = V^1 + V^2$ where $V^1(x) = \sum_{j=1}^m z_j |x - R_j|^{-1}$ and $1 \leq m < k$. Let $E_{\lambda}^{\text{TF},1}$ be the TF energy for the nuclei $1, \dots, m$ (with $U = \sum_{1 \leq i < j \leq m} z_i z_j |R_i - R_j|^{-1}$, of course) and let $E_{\lambda}^{\text{TF},2}$ be the same for the nuclei $m+1, \dots, k$. Given λ , let $\lambda_1 \geq 0$ and $\lambda_2 = \lambda - \lambda_1 \geq 0$ be chosen to minimize the sum of the energies of the separate molecules, i.e., $E_{\lambda_1}^{\text{TF},1} + E_{\lambda_2}^{\text{TF},2}$. (If $\lambda = Z = \sum_{j=1}^k z_j$ then by Theorem 3, $\lambda_1 = \sum_{j=1}^m z_j$.) Then

$$E_{\lambda}^{\text{TF}} \geq E_{\lambda_1}^{\text{TF},1} + E_{\lambda_2}^{\text{TF},2}. \quad (42)$$

Since the right side of Eq. (42) is the energy of two widely separated molecules, with the relative nuclear positions unchanged within each molecule, Theorem 6 says that the TF energy is unstable under *every* decomposition of the big molecule into smaller molecules. In particular, a molecule is unstable under decomposition into isolated atoms, and Theorem 9 is a simple consequence of this fact. One would suppose that if λ and the z_j are fixed, but the R_j are replaced by αR_j then

$$E_{\lambda}^{\text{TF}}(\{z_j, \alpha R_j\}_{j=1}^k) \text{ is monotone decreasing in } \alpha.$$

In other words, the "pressure" is always positive. This is an unproved *conjecture*, but it has been proved (Balázs, 1967) in the case $k=2$ and $z_1 = z_2$.

An interesting side remark is the following.

Theorem 7. If the TF energy (26), (27) is redefined by excluding the repulsion term U in (26), then the inequality in (42) is reversed.

Thus, the nuclear repulsion is essential for the no binding theorem 6.

Another useful fact for some further developments of the theory, especially the TF theory of solids and the TF theory of screening (Lieb and Simon, 1977) is the following lemma (also attributed to Teller), which is used to prove the main no binding theorem 6.

Lemma 8. Fix $\{R_j\}_{j=1}^k$ and fix $\mu \geq 0$ in the TF equation (28) but not $\{z_j\}_{j=1}^k$. (This means that as the z_j 's are varied λ will vary, but always $0 \leq \lambda \leq Z = \sum z_j$. If $\mu=0$ then $\lambda=Z$ always.) If $\{z_j^1\}_{j=1}^k$ and $\{z_j^2\}_{j=1}^k$ are two sets of z 's such that

$$z_j^1 \leq z_j^2 \text{ all } j, \text{ and } z_1^1 < z_1^2$$

and if λ_1 and λ_2 are the corresponding λ 's for the two sets, then for all x

$$\phi_{\lambda_1}^{\text{TF}}(x) \leq \phi_{\lambda_2}^{\text{TF}}(x)$$

and hence

$$\rho_{\lambda_1}^{\text{TF}}(x) \leq \rho_{\lambda_2}^{\text{TF}}(x).$$

There is strict inequality when $\mu=0$. In short, increasing some z_j increases the density everywhere, not just on the average.

The proof of Lemma 8 involves a beautifully simple potential theoretic argument which we cannot resist giving.

Proof of Lemma 8. We want to prove $\phi_1^{\text{TF}}(x) \leq \phi_2^{\text{TF}}(x)$ for all x and will content ourselves here with proving only $\leq$ when $\mu=0$. Let $B = \{x: \phi_1^{\text{TF}}(x) > \phi_2^{\text{TF}}(x)\}$. B is an open set and B does not contain any R_i for which $z_i^1 < z_i^2$ by the TF equation (29). Let $\psi(x) = \phi_1^{\text{TF}}(x) - \phi_2^{\text{TF}}(x)$. If $x \in B$ then $\psi(x) > 0$ and, by (28), $\rho_1^{\text{TF}}(x) \geq \rho_2^{\text{TF}}(x)$. For $x \in B$, $-(4\pi)^{-1} \Delta \psi(x) = \rho_2^{\text{TF}}(x) - \rho_1^{\text{TF}}(x) \leq 0$, so ψ is *subharmonic* on B [i.e., $\psi(x) \leq$ the average of ψ on any sphere contained in B and centered at x]. Hence ψ has its maximum on the boundary of B or at ∞ , at all of which points $\psi=0$. Therefore B is the empty set. ■

In the $\mu=0$ case it is easy to show how Theorem 6 follows from Lemma 8.

Proof of Theorem 6 when $\lambda = \sum_{j=1}^k z_j$. The proof when $\lambda < \sum z_j$ uses the same ideas but is more complicated. Since $\lambda = \sum z_j$, then $\lambda_1 = \sum_{j=1}^m z_j$, $\lambda_2 = \sum_{j=m+1}^k z_j$ and $\mu=0$ for all three systems. For $\alpha > 0$ let $f(\alpha) = E^{\text{TF}}(\alpha z_1, \dots, \alpha z_m; z_{m+1}, \dots, z_k; R_1, \dots, R_k) - E^{\text{TF}}(\alpha z_1, \dots, \alpha z_m; R_1, \dots, R_m) - E^{\text{TF}}(z_{m+1}, \dots, z_k; R_{m+1}, \dots, R_k)$, where the three E^{TF} are defined for neutral systems (i.e., $\mu=0$ for all α). The goal is to show that $f(1) \geq 0$. Since $f(0)=0$, it is enough to show that $df(\alpha)/d\alpha \geq 0$. From (26) and (27) it is true, and almost obvious, that

$$\begin{aligned} \frac{\partial E^{\text{TF}}}{\partial z_i} &= - \int \rho^{\text{TF}}(y) |y - R_i|^{-1} dy + \sum_{j \neq i} z_j |R_i - R_j|^{-1} \\ &= \lim_{x \rightarrow R_i} \phi^{\text{TF}}(x) - z_i |x - R_i|^{-1}. \end{aligned}$$

This is the TF version of the Feynman-Hellmann theorem; notice how the nuclear-nuclear repulsion comes in here. Thus,

$$\frac{df(\alpha)}{d\alpha} = \sum_{i=1}^m \lim_{x \rightarrow R_i} z_i \eta_{\alpha}(x),$$

where $\eta_{\alpha}(x) = \phi_1^{\text{TF}}(x) - \phi_2^{\text{TF}}(x)$ and ϕ_1^{TF} is the potential for $\{\alpha z_1, \dots, \alpha z_m, z_{m+1}, \dots, z_k; R_1, \dots, R_k\}$ and ϕ_2^{TF} is the potential for $\{\alpha z_1, \dots, \alpha z_m; R_1, \dots, R_m\}$. $\phi_1^{\text{TF}}(x) \geq \phi_2^{\text{TF}}(x)$ for all x by Lemma 8, and hence $\eta_{\alpha}(x) \geq 0$. ■

Theorem 6 has a natural application to the stability of matter problem. As will be shown in the next section, the TF energy (27) is, with suitably modified constants, a lower bound to the true quantum energy E_N^Q for all Z . By Theorem 3 (iv) and Theorem 6 we have the following theorem.

Theorem 9. Fix $\{z_j, R_j\}_{j=1}^k$ and let $Z = \sum_{j=1}^k z_j$. Then for all $\lambda \geq 0$

$$E_{\lambda}^{\text{TF}} \geq E_Z^{\text{TF}} \geq -(2.21) k^{2/3} (K^e)^{-1} \sum_{j=1}^k z_j^{7/3}. \quad (43)$$

The latter constant, 2.21, is obtained by numerically solving the TF equation for a single, neutral atom (J. F. Barnes, private communication). By scaling, Eq. (43)

holds for an choice of K^c in the definition (26) of $\mathcal{G}(\rho)$.

Theorem 9 is what will be needed for the H stability of matter because it says that the TF system is H stable, i.e., the energy is bounded below by a constant times the nuclear particle number (assuming that the z_j are bounded, of course).

Another application of Theorem 6 that will be needed is the following strange *inversion of the role of electrons and nuclei* in TF theory. It will enable us to give a lower bound to the true quantum-mechanical *electron-electron* repulsion. This theorem has nothing to do with quantum mechanics per se; it is really a theorem purely about electrostatics even though it is derived from the TF no binding theorem.

Theorem 10. Suppose that $x_1, \dots, x_N$ are any N distinct points in 3-space and define

$$V_X(y) = \sum_{j=1}^N |y - x_j|^{-1}. \quad (44)$$

Let $\gamma > 0$ and let $\rho(x)$ be any non-negative function such that $\int \rho(x) dx < \infty$ and $\int \rho(x)^{5/3} dx < \infty$. Then

$$\begin{aligned} \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} &\geq -\frac{1}{2} \int \int \rho(x) |x - y|^{-1} \rho(y) dx dy \\ &\quad + \int \rho(y) V_X(y) dy - (2.21)N/\gamma \\ &\quad - \gamma \int \rho(y)^{5/3} dy. \end{aligned} \quad (45)$$

Proof. Consider $\mathcal{G}(\rho)$ (26) with $q = 1$, $k = N$, K^c replaced by γ , $z_j \equiv 1$ and $R_j \equiv x_j$, $j = 1, \dots, N$. Let $\lambda = \int \rho(x) dx$. Then $\mathcal{G}(\rho) \geq E_\lambda^{\text{TF}}$ (by definition) and $E_\lambda^{\text{TF}} \geq -(2.21)N/\gamma$ by Theorem 9. The difference of the two sides in Eq. (45) is just $\mathcal{G}(\rho) + (2.21)N/\gamma$. ■

IV. THE STABILITY OF BULK MATTER

The various results of the last two sections can now be assembled to prove that the ground state energy (or infimum of the spectrum, if this not an eigenvalue) of H_N is bounded below by an extensive quantity, namely the total number of particles, independent of the nuclear locations $\{R_j\}$. This is called the H stability of matter to distinguish it from thermodynamic stability introduced in the next section. As explained before, the inclusion of the nuclear kinetic energy, as will be done in the next section, can only raise the energy.

The first proof of the N boundedness of the energy was given by Dyson and Lenard (Dyson and Lenard, 1967, Lenard and Dyson, 1968). Their proof is a remarkable analytic tour de force, but a chain of sufficiently many inequalities was used that they ended up with an estimate of something like -10^{14} Ry/particle. Using the results of the previous sections we will end up with -23 Ry/particle [see Eq. (56)].

We have in mind, of course, that the nuclear charges z_j , if they are not all the same, are bounded above by some fixed charge z .

Take any fermion $\psi(x_1, \dots, x_N; \sigma_1, \dots, \sigma_N)$ which is normalized and antisymmetric in the (x_i, σ_i) . Define the kinetic energy T_ψ and the single particle density ρ_ψ as in (16) and (17). We wish to compute a lower bound to

$$E_N^Q = \langle \psi, H_N \psi \rangle \quad (46)$$

with H_N being the N -particle Hamiltonian given in (23) and $\langle \psi, \psi \rangle = 1$.

For the third term on the right side of (23) Theorem 10 can be used with ρ taken to be ρ_ψ . Then, for any $\gamma > 0$

$$\begin{aligned} \left\langle \psi, \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} \psi \right\rangle &\geq \frac{1}{2} \int \int \rho_\psi(x) |x - y|^{-1} \rho_\psi(y) dx dy \\ &\quad - (2.21)N\gamma^{-1} - \gamma \int \rho_\psi(y)^{5/3} dy. \end{aligned} \quad (47)$$

Notice how the first and second terms on the right side of (45) combine to give $+\frac{1}{2}$ since

$$\left\langle \psi, \left\{ \int \rho_\psi(y) V_X(y) dy \right\} \psi \right\rangle = \int \int \rho_\psi(x) |x - y|^{-1} \rho_\psi(y) dx dy. \quad (48)$$

To control the kinetic energy in (23) Theorem 1 is used; the total result is then

$$\begin{aligned} E_N^Q &\geq \alpha \int \rho_\psi(x)^{5/3} dx - \int V(x) \rho_\psi(x) dx \\ &\quad + \frac{1}{2} \int \int \rho_\psi(x) |x - y|^{-1} \rho_\psi(y) dx dy \\ &\quad + U(\{z_j, R_j\}_{j=1}^N) - (2.21)N\gamma^{-1} \end{aligned} \quad (49)$$

with

$$\alpha = (4\pi q)^{-2/3} K^c - \gamma. \quad (50)$$

Restrict γ , which was arbitrary, so that $\alpha > 0$. Then, apart from the constant term $-(2.21)N\gamma^{-1}$, Eq. (49) is just $\mathcal{G}_\alpha(\rho_\psi)$, the Thomas-Fermi energy functional $\mathcal{G}$ applied to ρ_ψ , but with $q^{-2/3}K^c$ replaced by α . Since $\mathcal{G}_\alpha(\rho_\psi) \geq E_{\lambda, N}^{\text{TF}} \equiv \inf\{\mathcal{G}_\alpha(\rho) : \int \rho = N\}$ (by definition), and since the neutral case always has the lowest TF energy, as shown in Theorem 9, we have that

$$\mathcal{G}_\alpha(\rho_\psi) \geq -(2.21)\alpha^{-1} \sum_{j=1}^N z_j^{7/3}. \quad (51)$$

Thus we have proved the following:

Theorem 11. If ψ is a normalized, antisymmetric function of space and spin of N variables, and if there are q spin states associated with each particle then, for any $\gamma > 0$ such that α defined by Eq. (50) is positive,

$$\langle \psi, H_N \psi \rangle \geq -(2.21) \left\{ N\gamma^{-1} + \alpha^{-1} \sum_{j=1}^N z_j^{7/3} \right\}. \quad (52)$$

The optimum choice for γ is

$$\gamma = (4\pi q)^{-2/3} K^c \left[\left(\sum_{j=1}^N \frac{z_j^{7/3}}{N} \right)^{1/2} + 1 \right]^{-1}$$

in which case

$$E_N^Q \geq -(2.21) \frac{(4\pi q)^{2/3} N}{K^c} \left\{ 1 + \left[\sum_{j=1}^N \frac{z_j^{7/3}}{N} \right]^{1/2} \right\}^2. \quad (53)$$

This is the desired result, but some additional remarks are in order.

(1) Since $[1 + a^{1/2}]^2 \leq 2 + 2a$,

$$E_N^Q \geq -(4.42)(4\pi q)^{2/3}(K^c)^{-1} \left\{ N + \sum_{j=1}^k z_j^{7/3} \right\}. \quad (54)$$

Thus, provided the nuclear charges z_j are bounded above by some fixed z , E_N^Q is indeed bounded below by a constant times the total particle number $N + k$.

(2) Theorem 11 does not presuppose neutrality.

(3) For electrons, $q = 2$ and the prefactor in Eq. (53) is $-(2.08)N$. As remarked after Theorem 1, the unwanted constant $(4\pi)^{2/3}$ has been improved to $[4\pi/(1.83)]^{2/3}$. Using this, the prefactor becomes $-(1.39)N$. If $z_j = 1$ (hydrogen atoms) and $N = k$ (neutrality) then

$$E_N^Q \geq -(5.56)N = -(22.24)N R_y. \quad (55)$$

(4) The power law $z^{7/3}$ cannot be improved upon for large z because Theorem 5 asserts that the energy of an atom is indeed proportional to $z^{7/3}$ for large z .

(5) It is also possible to show that matter is indeed bulky. This will be proved for any ψ and any nuclear configuration (not just the minimum energy configuration) for which $E_N^Q \leq 0$. The minimizing nuclear configuration is, of course, included in this hypothesis. Then

$$0 \geq E_N^Q = \frac{1}{2} T_\psi + \langle \psi, H'_N \psi \rangle,$$

where H'_N is Eq. (23) but with a factor $\frac{1}{2}$ multiplying $\sum_{i=1}^N \Delta_i$. By Theorem 11, $\langle \psi, H'_N \psi \rangle \geq 2E_N$, where E_N is the right side of Eq. (53) (replace K^c by $K^c/2$ there). Therefore, the first important fact is that

$$T_\psi \leq 4|E_N|,$$

and this is bounded above by the total particle number.

Next, for any $p \geq 0$, it is easy to check that there is a $C_p > 0$ such that for any non-negative $\rho(x)$,

$$\begin{aligned} & \left\{ \int \rho(x)^{5/3} dx \right\}^{2/3} \int |x|^p \rho(x) dx \\ & \geq C_p \left\{ \int \rho(x) dx \right\}^{1+5p/6}. \end{aligned}$$

It is easy to find a minimizing ρ for this and to calculate C_p : $\rho(x)^{2/3} = 1 - |x|^p$ for $|x| \leq 1$; $\rho(x) = 0$, otherwise.

Since T_ψ satisfies Eq. (18) we have that

$$\left\langle \psi, \sum_{i=1}^N |x_i|^p \psi \right\rangle = \int |x|^p \rho_\psi(x) dx \geq C_p' N(N^{5/3}/|E_N|)^{p/2},$$

with $C_p' = C_p(K^c/4)^{p/2}(4\pi q)^{-p/3}$.

If it is assumed that $\sum z_j^{7/3}/N$ is bounded, and hence that $(N^{5/3}/|E_N|)^{p/2} > AN^{p/3}$ for some A , we reach the conclusion that the radius of the system is at least of the order $N^{1/3}$, as it should be.

The above analysis did not use any specific property of the Coulomb potential, such as the virial theorem. It is also applicable to the more general Hamiltonian $H_{n,k}$ in Eq. (58).

(6) The q dependence was purposely retained in Eq. (53) in order to say something about *bosons*. If $q = N$, then it is easy to see that the requirement of antisymmetry in ψ is *no restriction* at all. In this case then, one has simply

$$E_N^Q = \inf \text{spec } H_N$$

over all of $L^2(\mathbb{R}^3)^N$. Therefore

$$E_N^Q(\text{bosons}) \geq \frac{-(2.21)(4\pi)^{2/3}}{K^c} N^{5/3} \left\{ 1 + \left[\sum_{j=1}^k \frac{z_j^{7/3}}{N} \right]^{1/2} \right\}^2. \quad (56)$$

It was shown by Dyson and Lenard (Dyson and Lenard, 1967) that

$$E_N^Q(\text{bosons}) \geq -(\text{const})N^{5/3},$$

and by Dyson (Dyson, 1967) that

$$E_N^Q(\text{bosons}) \leq -(\text{const})N^{7/5}. \quad (57)$$

Proving Eq. (57) was not easy. Dyson had to construct a rather complicated variational function related to the type used in the BCS theory of superconductivity. Therefore *bosons are not stable* under the action of Coulomb forces, but the exact power law is not yet known. Dyson has conjectured that it is $\frac{7}{5}$.

In any event, the essential point has been made that Fermi statistics is essential for the stability of matter. The uncertainty principle for one particle, even in the strong form (5), together with intuitive notions that the electrostatic energy ought not to be very great, are insufficient for stability. The additional physical fact that is needed is that the kinetic energy increases as the $\frac{5}{3}$ power of the fermion density.

V. THE THERMODYNAMIC LIMIT

Having established that E_N^Q is bounded below by the total particle number, the next question to consider is whether, under appropriate conditions, E_N^Q/N has a limit as $N \rightarrow \infty$, as expected. More generally, the same question can be asked about the free energy per particle when the temperature is not zero and the particles are confined to a box.

It should be appreciated that the difficulty in obtaining the lower bound to E_N^Q came almost entirely from the r^{-1} short range singularity of the Coulomb potential. Other potentials, such as the Yukawa potential, with the same singularity would present the same difficulty which would be resolved in the same way. The singularity was tamed by the $\rho^{5/3}$ behavior of the fermion kinetic energy.

The difficulty for the thermodynamic limit is different. It is caused by the long range r^{-1} behavior of the Coulomb potential. In other words, we are faced with the problem of explosion rather than implosion. Normally, a potential that falls off with distance more slowly than $r^{-3-\epsilon}$ for some $\epsilon > 0$ does not have a thermodynamic limit. Because the charges have different signs, however, there is hope that a cancellation at large distances may occur.

An additional physical hypothesis will be needed, namely *neutrality*. To appreciate the importance of neutrality consider the case that the electrons have positive, instead of negative charge. Then $E_N^Q > 0$ because every term in Eq. (23) would be positive. While the H -stability question is trivial in this case, the thermodynamic limit is not. If the particles are constrained to be in a domain Ω whose volume $|\Omega|$ is proportional to N , the particles will repel each other so strongly that they will all go to the boundary of Ω in order to mini-

mize the electrostatic energy. The minimum electrostatic energy will be of the order $+N^2|\Omega|^{-1/3} \sim +N^{5/3}$. Hence no thermodynamic limit will exist.

When the system is neutral, however, the energy can be expected to be extensive, i.e., $O(N)$. For this to be so, different parts of the system far from each other must be approximately independent, despite the long range nature of the Coulomb force. The fundamental physical, or rather electrostatic, fact that underlies this is *screening*; the distribution of the particles must be sufficiently neutral and isotropic locally so that according to Newton's theorem (13 below) the electric potential far away will be zero. The problem is to express this idea in precise mathematical form.

We begin by defining the Hamiltonian for the *entire* system consisting of k nuclei, each of charge z and mass M , and n electrons ($\hbar^2/2=1$, $m=1$, $|e|=1$):

$$H_{n,k} = - \sum_{j=1}^n \Delta_j - \frac{1}{M} \sum_{j=n+1}^{n+k} \Delta_j - z \sum_{i=1}^n \sum_{j=n+1}^{n+k} |x_i - y_j|^{-1} + \sum_{1 \leq i < j \leq n} |x_i - x_j|^{-1} + z^2 \sum_{n+1 \leq i < j \leq n+k} |y_i - y_j|^{-1}. \quad (58)$$

The first and second terms in Eq. (58) are, respectively, the kinetic energies of the electrons and the nuclei. The last three terms are, respectively, the electron-nuclear, electron-electron, and nuclear-nuclear Coulomb interactions. The electron coordinates are x_i and the nuclear coordinates are y_i . The electrons are fermions with spin $\frac{1}{2}$; the nuclei may be either bosons or fermions.

The basic neutrality hypotheses is that n and k are related by

$$n = kz. \quad (59)$$

It is assumed that z is rational.

The thermodynamic limit to be discussed here can be proved under more general assumptions, i.e., we can have several kinds of negative particles (but they must all be fermions in order that the basic stability estimate of Sec. IV holds) and several kinds of nuclei with different statistics, charges, and masses. Neutrality must always hold, however. Short range forces and hard cores, in addition to the Coulomb forces, can also be included with a considerable sacrifice in simplicity of the proof. See (Lieb and Lebowitz, 1972).

$H_{n,k}$ acts on square integrable functions of $n+k$ variables (and spin as well). To complete the definition of $H_{n,k}$ we must specify boundary conditions: choose a domain Ω (an open set, which need not be connected) and require that $\psi=0$ if x_i or y_i are on the boundary of Ω .

For each non-negative integer j , choose an n_j and a corresponding k_j determined by Eq. (59), and choose a domain Ω_j . The symbol N_j will henceforth stand for the pair (n_j, k_j) and

$$|N_j| \equiv n_j + k_j.$$

We require that the densities

$$\rho_j \equiv |N_j| |\Omega_j|^{-1} \quad (60)$$

be such that

$$\lim_{j \rightarrow \infty} \rho_j = \rho. \quad (61)$$

ρ is then the density in the thermodynamic limit. Here we shall choose the Ω_j to be a sequence of *balls* of radii R_j and shall denote them by B_j .

It can be shown that the *same* thermodynamic limit for the energy and free energy holds for any sequence N_j, Ω_j and depends *only* on the limiting ρ and β , and not on the "shape" of the Ω_j , provided the Ω_j go to infinity in some reasonable way.

The basic quantity of interest is the *canonical partition function*

$$Z(N, \Omega, \beta) = \text{Tr} \exp(-\beta H_{n,k}), \quad (62)$$

where the trace is on $L^2(\Omega)^{|N|}$ and $\beta=1/T$, T being the temperature in units in which Boltzmann's constant is unity.

The *free energy per unit volume* is

$$F(N, \Omega, \beta) = -\beta^{-1} \ln Z(N, \Omega, \beta) / |\Omega| \quad (63)$$

and the problem is to show that with

$$F_j = F(N_j, \Omega_j, \beta), \quad (64)$$

then

$$\lim_{j \rightarrow \infty} F_j \equiv F(\rho, \beta) \quad (65)$$

exists. A similar problem is to show that

$$E(N, \Omega) \equiv |\Omega|^{-1} \inf_{\psi} \langle \psi, H_{n,k} \psi \rangle / \langle \psi, \psi \rangle, \quad (66)$$

the *ground state energy per unit volume*, has a limit

$$e(\rho) = \lim_{j \rightarrow \infty} E_j, \quad (67)$$

where

$$E_j = E(N_j, \Omega_j).$$

The proof we will give for the limit $F(\rho, \beta)$ will hold equally well for $e(\rho)$ because E_j can be substituted for F_j in all statements.

The basic strategy consists of two parts. The easiest part is to show that F_j is bounded below. We already know this for E_j by the results of Sec. IV. The second step is to show that in some sense the sequence F_j is decreasing. This will then imply the existence of a limit.

Theorem 12. Given N , Ω , and β there exists a constant C depending only on $\rho = |N|/|\Omega|$ and β such that

$$F(N, \Omega, \beta) \geq C. \quad (68)$$

Proof. Write $H_{n,k} = H_A + H_B$, where

$$H_A = -\frac{1}{2} \left\{ \sum_{i=1}^n \Delta_i + \frac{1}{M} \sum_{j=n+1}^{n+k} \Delta_j \right\}$$

is half the kinetic energy. Then $H_B \geq b|N|$, with b depending only on z , by the results of Sec. IV (increasing the mass by a factor of 2 in H_B only changes the constant b). Hence $Z(N, \Omega, \beta) \leq \exp(-b\beta|N|) \text{Tr} \exp(-\beta H_A)$. However, $\text{Tr} \exp(-\beta H_A)$ is the partition function of an ideal gas and it is known by explicit computation that it is bounded above by $e^{b_d|N|}$ with d depending only on $\rho = |N|/|\Omega|$ and β . Thus

$$F(N, \Omega, \beta) \geq (b-d)\rho. \quad \blacksquare$$

For the second step, two elementary but basic in-

equalities used in the general theory of the thermodynamic limit are needed and they will be described next.

A. Domain partition inequality: Given a domain Ω and the particle numbers $N = (n, k)$, let π be a partition of Ω into l disjoint domains $\Omega^1, \dots, \Omega^l$. Likewise N is partitioned into l integral parts (some of which may be zero):

$$N = N^1 + \dots + N^l.$$

Then for any such partition, π , of Ω and N

$$Z(N, \Omega, \beta) = \text{Tr} \exp(-\beta H_{n,k}) \geq \text{Tr}^\pi \exp(-\beta H_N^\pi). \quad (69)$$

Here Tr^π means trace over

$$\mathfrak{H}^\pi = L^2(\Omega^1)^{|N^1|} \otimes \dots \otimes L^2(\Omega^l)^{|N^l|},$$

and H_N^π is defined as in (58) but with Dirichlet ($\psi=0$) boundary conditions for the N^i particles on the boundary of Ω^i (for $i=1, \dots, l$).

Simply stated, the first N^1 particles are confined to Ω^1 , the second N^2 to Ω^2 , etc. The interaction among the particles in different domains is still present in H_N^π . Equation (69) can be proved by the Peierls-Bogoliubov variational principle for $\text{Tr} e^X$. Alternatively, (69) can be viewed simply as the statement that the insertion of a hard wall, infinite potential on the boundaries of the Ω^i only decreases Z ; the further restriction of a definite particle number to each Ω^i further reduces Z because it means that the trace is then over only the H_N^π -invariant subspace, $\mathfrak{H}^\pi$, of the full Hilbert space.

B. Inequality for the interdomain interaction: The second inequality is another consequence of the convexity of $A \rightarrow \text{Tr} e^A$ (Peierls-Bogoliubov inequality):

$$\text{Tr} e^{A+B} \geq \text{Tr} e^A \exp \langle B \rangle, \quad (70)$$

where

$$\langle B \rangle \equiv \text{Tr} B e^A / \text{Tr} e^A. \quad (71)$$

Some technical conditions are needed here, but Eqs. (70) and (71) will hold in our application.

To exploit (70), first make the same partition π as in inequality A and then write

$$H_N^\pi = H_0 + W(X), \quad (72)$$

$$H_0 = H^1 + \dots + H^l, \quad (73)$$

with H^i being that part of the total Hamiltonian (58) involving only the N^i particles in Ω^i , and H^i is defined with the stated Dirichlet boundary conditions on the boundary of Ω^i . $W(X)$, with X standing for all the coordinates, is the interdomain Coulomb interaction. In other words, $W(X)$ is that part of the last three terms on the right side of (58) which involves coordinates in different blocks of the partition π . Technically, W is a small perturbation of H_0 .

With

$$A = -\beta H_0 \text{ and } B = -\beta W \quad (74)$$

in (70), we must calculate $\langle W \rangle$. Since $e^A = e^{-\beta H_0}$ is a simple tensor product of operators on each $L^2(\Omega^i)^{|N^i|}$, W is merely the average interdomain Coulomb energy in a canonical ensemble in which the Coulomb interaction is present in each subdomain but the l domains are independent of each other. In other words, let $q^i(x)$,

$x \in \Omega^i$ denote the average charge density in Ω^i for this ensemble of independent domains, namely

$$q^i(x) = \sum_{j=1}^{|N^i|} q_j \int_{|\Omega^i|^{-1}} \exp(-\beta H^i)(X^i, X^i) \hat{d}x_j / Z(N^i, \Omega^i, \beta) \quad (75)$$

with the following notation: X^i stands for the coordinates of the $|N^i|$ particles in Ω^i , $\hat{d}x_j$ means integration over all these coordinates (in Ω^i) with the exception of x_j , and x_j is set equal to x ; q_j is the charge (-1 or $+z$) of the j th particle; $\exp(-\beta H^i)(X^i, Y^i)$ is a kernel (x -space representation) for $\exp(-\beta H^i)$. $q^i(x)$ vanishes if $x \notin \Omega^i$.

With the definitions (75) one has that

$$\langle W \rangle = \sum_{i < j} \int_{\Omega_i} \int_{\Omega_j} q^i(x) q^j(y) |x - y|^{-1} dx dy. \quad (76)$$

Equation (70), together with (76) and (74), is the desired inequality for the interdomain interaction. It is quite general in that an analogous inequality holds for arbitrary two-body potentials. Neither specific properties of the Coulomb potential nor neutrality was used.

Now we come to the crucial point at which screening is brought in. The following venerable result from the *Principia Mathematica* is essential.

Theorem 13 (Newton). Let $\rho(x)$ be an integrable function on 3-space such that $\rho(x) = \rho(y)$ if $|x| = |y|$ (isotropy) and $\rho(x) = 0$ if $|x| > R$ for some $R > 0$. Let

$$\phi(x) = \int \rho(y) |x - y|^{-1} dy \quad (77)$$

be the Coulomb potential generated by ρ . Then if $|x| \geq R$

$$\phi(x) = |x|^{-1} \int \rho(y) dy. \quad (78)$$

The important point is that an isotropic, neutral charge distribution generates zero potential outside its support, irrespective of how the charge is distributed radially.

Suppose that N^i is neutral, i.e., the electron number $= z$ times the nucleon number for each subdomain in Ω . Suppose also that the subdomain Ω^i is a ball of radius R^i centered at a^i . Then since H^i is rotation invariant, $q^i(x) = q^i(y)$ if $|x - a^i| = |y - a^i|$, $\int q^i(x) dx = 0$ (by neutrality) and $q^i(x) = 0$ if $|x - a^i| > R^i$. Then, by Theorem 13, every term in Eq. (76) involving q^i vanishes, because when $j \neq i$, $q^j(y) = 0$ if $|y - a^j| < R^j$ since Ω^j is disjoint from Ω^i . Consequently the average interdomain interaction, $\langle W \rangle$, vanishes.

In the decomposition, π , of Ω into $\Omega^1, \dots, \Omega^l$ and N into $N^1, \dots, N^l$ we will arrange matters such that

- (i) $\Omega^1, \dots, \Omega^{l-1}$ are balls,
- (ii) $N^1, \dots, N^{l-1}$ are neutral,
- (iii) $N^l = 0$.

Then $\langle W \rangle = 0$ and, using Eqs. (69) and (70)

$$\begin{aligned} Z(N, \Omega, \beta) &\geq \text{Tr}^\pi \exp(-\beta H_N^\pi) \geq \prod_{i=1}^{l-1} Z(N^i, \Omega^i, \beta) e^{-\beta \langle W \rangle} \\ &= \prod_{i=1}^{l-1} Z(N^i, \Omega^i, \beta). \end{aligned} \quad (79)$$

In addition to (i), (ii), (iii) it will also be necessary to arrange matters such that when Ω is a ball B_K in the chosen sequence of domains, then the subdomains $\Omega^1, \dots, \Omega^{l-1}$ in the partition of B_K are also smaller balls in the same sequence. With these requirements in mind the *standard sequence*, which depends on the limiting density ρ , is defined as follows:

- (1) Choose $\rho > 0$.
- (2) Choose any N_0 satisfying the neutrality condition (59).
- (3) Choose R_0 such that

$$28(4\pi/3)\rho R_0^3 = |N_0|. \quad (80)$$

- (4) For $j \geq 1$ let

$$R_j = (28)^j R_0, \quad (81)$$

$$N_j = (28)^{3j-1} N_0$$

be the radius of the ball B_j and the particle number in that ball.

It will be noted that the density in all the balls except the first is

$$\rho_j = \rho, \quad j \geq 1, \quad (82)$$

while the density in the smallest ball is much bigger:

$$\rho_0 = 28\rho. \quad (83)$$

This has been done so that when a ball B_K , $K \geq 1$ is packed with smaller balls in the manner to be described below, the density in each ball will come out right; the higher density in B_0 compensates for the portion of B_K not covered by smaller balls. The radii increase geometrically, namely by a factor of 28.

The number 28 may be surprising until it is realized that the objective is to be able to pack B_K with balls of type B_{K-1}, B_{K-2} , etc., in such a way that as much as possible of B_K is covered and *also* that very little of B_K is covered by very small balls. If the ratio of radii were too close to unity, then the packing of B_K would be inefficient from this point of view. In short, if the number 28 is replaced by a much smaller number the analog of the following basic geometric theorem will not be true.

Theorem 14 (Cheese theorem). For j a positive integer define the integer $m_j \equiv (27)^{j-1}(28)^{2j}$. Then for each positive integer $K \geq 1$ it is possible to pack the ball B_K of radius R_K (given by 81) with

$$\bigcup_{j=0}^{K-1} (m_{K-j} \text{ balls of radius } R_j).$$

"Pack" means that all the balls in the union are disjoint.

We will not give a proof of Theorem 14 here, but note that it entails showing that m_1 balls of radius R_{K-1} can be packed in B_K in a cubic array, then that m_2 balls of radius R_{K-2} can be packed in a cubic array in the interstitial region, etc.

Theorem 14 states that B_K can be packed with $(28)^2$ balls of type B_{K-1} , $(27)(28)^4$ balls of type B_{K-2} , etc. If f_{K-j} is the fraction of the volume of B_K occupied by all the balls of radius R_j in the packing, then

$$f_j = m_j (R_{K-j}/R_K)^3 = \frac{1}{27} \gamma^j \quad (84)$$

with

$$\gamma = \frac{27}{28} < 1. \quad (85)$$

The packing is *asymptotically complete* in the sense that

$$\lim_{K \rightarrow \infty} \sum_{j=0}^{K-1} f_{K-j} = \frac{1}{27} \sum_{j=1}^{\infty} \gamma^j = 1. \quad (86)$$

It is also "geometrically rapid" because the fraction of $|B_K|$ that is uncovered is

$$\sum_{j=K+1}^{\infty} f_j = \gamma^K. \quad (87)$$

The necessary ingredients having been assembled, we can now prove the following theorem.

Theorem 15. Given ρ and $\beta > 0$, the thermodynamic limits $F(\rho, \beta)$ and $e(\rho)$ (65, 67) exist for the sequence of balls and particle numbers specified by (80) and (81).

Proof. Let F_K given by Eq. (64) be the free energy per unit volume for the ball B_K with N_K particles in it. For $K \geq 1$, partition B_K into disjoint domains $\Omega^1, \dots, \Omega^l$, where the Ω^i for $i = 1, \dots, l-1$ designate the smaller balls referred to in Theorem 14, and Ω^l (which is the "cheese" after the holes have been removed) is the remainder of B_K . The smaller balls are copies of B_j , $0 \leq j \leq K-1$; in each of these place N_j particles according to (81). $N^l = 0$. The total particle number in B_K is then

$$\sum_{j=0}^{K-1} N_j m_{K-j} = N_0 \{ (27)^{K-1} (28)^{2K} + \sum_{j=1}^{K-1} (28)^{3j-1} (27)^{K-j-1} (28)^{2K-2j} \} \\ = N_0 (28)^{3K-1} = N_K$$

as it should be.

Use the basic inequality (79); $\langle W \rangle = 0$ since all the smaller balls are neutral and Ω^l contains no particles. Thus, taking logarithms and dividing by $|B_K|$, we have for $K \geq 1$ that

$$F_K \leq \sum_{j=0}^{K-1} F_j f_{K-j} \quad (88)$$

with $f_j = \gamma^j/27$ and $\gamma = \frac{27}{28}$. This inequality can be rewritten as

$$F_K = \sum_{j=0}^{K-1} \frac{F_j \gamma^{K-j}}{27} - d_K \quad (89)$$

with $d_K \geq 0$. Equation (89) is a *renewal equation* which can be solved explicitly by inspection:

$$F_K = -\gamma d_K - \sum_{j=1}^K \frac{d_j}{28} + \frac{F_0}{28}. \quad (90)$$

We now use the first step, Theorem 13, on the boundedness of F_K . Since $F_K \geq C$, $\sum_{j=1}^{\infty} d_j$ must be finite, for otherwise (90) would say that $F_K \rightarrow -\infty$. The convergence of the sum implies that $d_K \rightarrow 0$ as $K \rightarrow \infty$. Hence the limit exists; specifically

$$F = \lim_{K \rightarrow \infty} F_K = -\sum_{j=1}^{\infty} \frac{d_j}{28} + \frac{F_0}{28}. \quad \blacksquare \quad (91)$$

Theorem 15 is the desired goal, namely the existence of the thermodynamic limit for the free energy (or ground state energy) per unit volume. There are, how-

ever, some additional points that deserve comment.

(A) For each given limiting density ρ , a particular sequence of domains, namely balls, and particle numbers was used. It can be shown that the same limit is reached for general domains, with some mild conditions on their shape including, of course, balls of different radii than that used here. The argument involves packing the given domains with balls of the standard sequence and vice versa. The proof is tedious, but standard, and can be found in (Lieb and Lebowitz, 1972).

(B) Here we have considered the thermodynamic limit for real matter, in which all the particles are mobile. There are, however, other models of some physical interest. One is *jellium* in which the positive nuclei are replaced by a *fixed, uniform background* of positive charge. With the aid of an additional trick the thermodynamic limit can also be proved for this model (Lieb and Narnhofer, 1975). Another, more important model is one in which the nuclei are *fixed point charges* arranged periodically in a lattice. This is the model of solid state physics. Unfortunately, local rotation invariance is lost and Newton's Theorem 13 cannot be used. This problem is still open and its solution will require a deeper insight into screening.

(C) An absolute physical requirement for $\beta F(\rho, \beta)$, as a function of $\beta = 1/T$, is that it be *concave*. This is equivalent to the fact that the specific heat is non-negative since (specific heat) $= -\beta^2 \partial^2 \beta F(\rho, \beta) / \partial \beta^2$. Fortunately it is true. From the definitions (57), (58) we see that $\ln Z(N, \Omega, \beta)$ is convex in β for every finite system and hence $\beta F(N, \Omega, \beta)$ is concave. Since the limit of a sequence of concave functions is always concave, the limit $\beta F(\rho, \beta)$ is concave in β .

(D) Another absolute requirement is that $F(\rho, \beta)$ be *convex* as a function of ρ . This is called *thermodynamic stability* as distinct from the lower bound H stability of the previous sections. It is equivalent to the fact that the compressibility is non-negative, since (compressibility) $^{-1} = \partial P / \partial \rho = \rho \partial^2 F(\rho, \beta) / \partial \rho^2$. Frequently, in approximate theories (e.g., van der Waals' theory of the vapor-liquid transition, some field theories, or some theories of magnetic systems in which the magnetization per unit volume plays the role of ρ), one introduces an F with a double bump. Such an F is nonphysical and never should arise in an exact theory.

For a finite system, F is defined only for integral N , and hence not for all real ρ . It can be defined for all ρ by linear interpolation, for example, but even so it can neither be expected to be, nor is it generally, convex, except in the limit. The idea behind the following proof is standard.

Theorem 16. The limit function $F(\rho, \beta)$ is a convex function of ρ for each fixed β . $E(\rho)$ is also a convex function of ρ .

Proof: This means that for $\rho = \lambda \rho^1 + (1 - \lambda) \rho^2$, $0 \leq \lambda \leq 1$,

$$F(\rho, \beta) \leq \lambda F(\rho^1, \beta) + (1 - \lambda) F(\rho^2, \beta) \quad (92)$$

and similarly for $E(\rho)$. As F is bounded above on bounded ρ intervals (this can be proved by a simple variational calculation), it is sufficient to prove (92) when $\lambda = \frac{1}{2}$. To avoid technicalities (which can be sup-

plied) and concentrate on the main idea, we shall here prove (92) when ρ^2 and ρ^1 are rationally related: $a\rho^1 = b\rho^2$, a and b positive integers. Choose any neutral particle number M and define a sequence of balls B_j with radii as given in (81) and with $28(4\pi/3)\rho R_0^3 = (a+b)|M|$. For the ρ system take $N_0 = (a+b)M$, $N_j = (28)^{j+1}N_0$, $j \geq 1$. For the ρ^1 (respectively, ρ^2) system take $N_0^1 = 2bM$, $N_j^1 = (28)^{j+1}N_0^1$ [respectively, $N_0^2 = 2aM$, $N_j^2 = (28)^{j+1}N_0^2$]. Consider the ρ system. In the canonical partition π of B_K into smaller balls (Theorem 14) note that the number of balls B_j is m_{K-j} and this number is *even*. In half of these balls place N_j^1 particles and in the other half place N_j^2 particles, $0 \leq j \leq K-1$. Then in place of (88) we get

$$F_K(\rho) \leq \frac{1}{2} \sum_{j=0}^{K-1} f_{K-j} [F_j(\rho^1) + F_j(\rho^2)] \quad (93)$$

in an obvious notation. Inserting (89) on the right side of (93),

$$F_K(\rho) \leq \frac{1}{2} [F_K(\rho^1) + F_K(\rho^2)] + \frac{1}{2} (d_K^2 + d_K^2). \quad (94)$$

Since $\lim_{K \rightarrow \infty} d_K^2 = 0$, we can take the limit $K \rightarrow \infty$ in Eq. (94) and obtain (92). ■

(E) The convexity in ρ^1 and concavity in β of $F(\rho, \beta)$ has another important consequence. Since F is bounded below (Theorem 13) and bounded above (by a simple variational argument) on bounded sets in the (ρ, β) plane, the convexity/concavity implies that it is *jointly continuous* in (ρ, β) . This, together with the monotonicity in K of $F_K + \gamma d_K$ (see (90)), implies by a standard argument using Dini's theorem that *the thermodynamic limit is uniform* on bounded (ρ, β) sets. This uniformity is sometimes overlooked as a basic desideratum of the thermodynamic limit. Without it one would have to fix ρ and β *precisely* in taking the limit—an impossible task experimentally. With it, it is sufficient to have merely an increasing sequence of systems such that $\rho_j \rightarrow \rho$ and $\beta_j \rightarrow \beta$. The same result holds for $e(\rho)$.

(F) An application of the uniformity of the limit for $e(\rho)$ is the following. Instead of confining the particles to a box (Dirichlet boundary condition for $H_{n,K}$) one could consider $H_{n,K}$ defined on all of $L^2(\mathbb{R}^3)|N|$, i.e., no confinement at all. In this case

$$E_N^Q \equiv \inf_{\psi} \langle \psi, H_{n,K} \psi \rangle / \langle \psi, \psi \rangle$$

is just the ground state energy of a neutral molecule and it is expected that $E_N^Q/|N|$ has a limit. Indeed, this limit exists and it is simply

$$\lim_{n \rightarrow \infty} E_N^Q/|N| = \lim_{\rho \rightarrow 0} \rho^{-1} e(\rho).$$

There is no analog of this for $F(\rho, \beta)$ because removing the box would cause the partition function to be infinite even for a finite system.

(G) The ensemble used here is the *canonical* ensemble. It is possible to define and prove the existence of the thermodynamic limit for the *microcanonical* and *grand canonical* ensembles and to show that all three ensembles are equivalent (i.e., that they yield the same values for all thermodynamic quantities, such as the pressure). (See Lieb and Lebowitz, 1972.)

(H) Charge neutrality was essentially for taming the long range Coulomb force. What happens if the system is *not neutral*? To answer this let N_j, Ω_j be a sequence

of pairs of particle numbers and domains, but without (59) being satisfied. Let $Q_j = zk_j - n_j$ be the net charge, $\rho_j = |N_j|/|\Omega_j|$ as before, and $\rho_j \rightarrow \rho$. One expects that if (i) $Q_j |\Omega_j|^{-2/3} \rightarrow 0$ then the same limit $F(\rho, \beta)$ is achieved as if $Q_j = 0$.

On the other hand, if

(ii) $Q_j |\Omega_j|^{-2/3} \rightarrow \infty$ then there is no limit for $F(N_j, \Omega_j, \beta)$. More precisely $F(N_j, \Omega_j, \beta) \rightarrow \infty$ because the minimum electrostatic energy is too great. Both of these expectations can be proved to be correct.

The interesting case is if

(iii) $\lim_{j \rightarrow \infty} Q_j |\Omega_j|^{-2/3} = \sigma$ exists. Then one expects a shape dependent limit to exist as follows. Assume that the Ω_j are geometrically similar, i.e., $\Omega_j = \lambda_j \Omega_0$ with $|\Omega_0| = 1$ and $|N_j| \lambda_j^{-3} = \rho_j$ with $\rho_j \rightarrow \rho$. Let C be the electrostatic capacity of Ω_0 ; it depends upon the shape of Ω_0 . The capacity of Ω_j is then $C_j = C \lambda_j$. From elementary electrostatics theory the expectation is that

$$\lim_{j \rightarrow \infty} F(N_j, \Omega_j, \beta) = F(\rho, \beta) + \sigma^2/2C. \quad (95)$$

Note that $(Q_j^2/2C_j) |\Omega_j|^{-1} \rightarrow \sigma^2/2C$.

Equation (95) can be proved for ellipsoids and balls. The proof is as complicated as the result is simple. With work, the proof could probably be pushed through for other domains Ω_0 with smooth boundaries.

The result (95) is amazing and shows how special the Coulomb force is. It says that the surplus charge Q_j goes to a thin layer near the surface. There, only its electrostatic energy, which overwhelms its kinetic energy, is significant. The bulk of Ω_j is neutral and uninfluenced by the surface layer because the latter generates a constant potential inside the bulk. It is seldom that one has two strongly interacting subsystems and that the final result has no cross terms, as in Eq. (95).

(I) There might be a temptation, which should be avoided, to suppose that the thermodynamic limit describes a single phase system of uniform density. The temptation arises from the construction in the proof of Theorem 15 in which a large domain B_R is partitioned into smaller domains having essentially constant density. Several phases can be present inside a large domain. Indeed, if β is very large a solid is expected to form, and if the average density, ρ , is smaller than the equilibrium density, ρ_s , of the solid a dilute gas phase will also be present. The location of the solid inside the larger domain will be indeterminate.

From this point of view, there is an amusing, although expected, aspect to the theorem given in Eq. (95). Suppose that β is very large and that $\rho < \rho_s$. Suppose, also, that a surplus charge $Q = \sigma V^{2/3}$ is present, where V is the volume of the container. In equilibrium, the surplus charge will never be bound to the surface of the solid, for that would give rise to a larger free energy than in (95).

(J) The inequality (53) of Sec. IV, together with known facts about the ideal gas, permit one to derive upper and lower bounds to the free energy and pressure for any neutral mixture of electrons and various nuclei. These bounds are absolutely rigorous and involve no approximation whatsoever (beyond the assumption of nonrelativistic Schrödinger mechanics with purely Coulomb forces).

If one has bounds on the free energy per unit volume

$$F^L(\rho, \beta) \leq F(\rho, \beta) \leq F^U(\rho, \beta), \quad (96)$$

then since the pressure P is equal to $-F + \rho \partial F / \partial \rho$, and since F is convex in ρ , one has that

$$\begin{aligned} P &\leq -F + \rho \min_{\epsilon > 0} \epsilon^{-1} \{F(\rho + \epsilon, \beta) - F(\rho, \beta)\}, \\ P &\geq -F + \rho \max_{\epsilon > 0} \epsilon^{-1} \{F(\rho, \beta) - F(\rho - \epsilon, \beta)\}. \end{aligned} \quad (97)$$

Inserting (96) into (97) yields bounds on P .

Equation (96) comes from bounds on Z [see Eq. (63)]. Using (70) and $\rho = \rho_{\text{nuc}} + \rho_{\text{el}}$

$$F^U(\rho, \beta) = F_{\text{el}}^0(\rho_{\text{el}}, \beta) + F_{\text{nuc}}^0(\rho_{\text{nuc}}, \beta) + \langle W \rangle / |\Omega|, \quad (98)$$

where F^0 is the ideal gas free energy, and $\langle W \rangle / |\Omega|$ is the average total Coulomb energy per unit volume in the ideal gas state. This can easily be computed in terms of exchange integrals. To obtain F^L , choose $0 < \gamma < 1$ and write $H_{\gamma, h} = (1 - \gamma)T_{\text{el}} + T_{\text{nuc}} + h(\gamma)$, where T is the kinetic energy operator, and $h(\gamma) = \gamma T_{\text{el}} + W$. $h(\gamma)$ is bounded below by $A/\gamma \equiv [\text{right side of Eq. (53)}]/\gamma$. Thus

$$Z \leq \exp[-\beta h(\gamma)] \text{Tr} \exp[-\beta \{(1 - \gamma)T_{\text{el}} + T_{\text{nuc}}\}]$$

and

$$\begin{aligned} F^L = F_{\text{nuc}}^0(\rho_{\text{nuc}}, \beta) + \max_{0 < \gamma < 1} \{ &(1 - \gamma)F_{\text{el}}^0(\rho_{\text{el}}, (1 - \gamma)\beta) \\ &+ \gamma^{-1}A/|\Omega| \}. \end{aligned} \quad (99)$$

A numerical evaluation of these bounds will be presented elsewhere.

As a final remark, the existence of the thermodynamic limit (and hence the existence of intensive thermodynamic variables such as the pressure) does not establish the existence of a unique thermodynamic state. In other words, it has not been shown that correlation functions, which always exist for finite systems, have unique limits as the volume goes to infinity. Indeed, unique limits might not exist if several phases are present. For well behaved potentials there are techniques available for proving that a state exists when the density is small, but these techniques do not work for the long range Coulomb potential. Probably the next chapter to be written in this subject will consist of a proof that correlation functions are well defined in the thermodynamic limit when ρ or β is small.

REFERENCES

- Balázs, N., 1967, "Formation of stable molecules within the statistical theory of atoms," Phys. Rev. **156**, 42-47.
- Barnes, J. F., 1975, private communication.
- Birman, M. S., 1961, Mat. Sb. **55** (97), 125-174 ["The spectrum of singular boundary value problems," Am. Math. Soc. Transl. Ser. 2 **53**, 23-80 (1966)].
- Dirac, P. A. M., 1930, "Note on exchange phenomena in the Thomas atom," Proc. Camb. Phil. Soc. **26**, 376-385.
- Dyson, F. J., 1967, "Ground-state energy of a finite system of charged particles," J. Math. Phys. **8**, 1538-1545.
- Dyson, F. J., and A. Lenard, 1967, "Stability of matter. I," J. Math. Phys. **8**, 423-434.
- Fermi, E., 1927, "Un metodo statistico per la determinazione di alcune proprietà dell'atome," Rend. Acad. Naz. Lincei **6**,

- 602–607.
- Fock, V., 1930, “Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems,” *Z. Phys.* **61**, 126–148; see also V. Fock, 1930, “Selfconsistent field” mit austausch für Natrium,” *Phys.* **62**, 795–805.
- Gombás, P., 1949, *Die statistischen Theorie des Atomes und ihre Anwendungen* (Springer Verlag, Berlin).
- Hartree, D. R., 1927–1928, “The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods,” *Proc. Camb. Phil. Soc.* **24**, 89–110.
- Heisenberg, W., 1927, “Über den anschaulichen Inhalt der quanten-theoretischen Kinematik und Mechanik,” *Z. Phys.* **43**, 172–198.
- Jeans, J. H., 1915, *The Mathematical Theory of Electricity and Magnetism* (Cambridge University, Cambridge) 3rd ed., p. 168.
- Kato, T., “Fundamental properties of Hamiltonian operators of Schrödinger type,” *Trans. Am. Math. Soc.* **70**, 195–211.
- Kirzhnits, D. A., 1957, *J. Exptl. Theoret. Phys. (USSR)* **32**, 115–123 [Engl. transl. “Quantum corrections to the Thomas–Fermi equation,” *Sov. Phys. JETP* **5**, 64–71 (1957)].
- Kompaneets, A. S., and E. S. Pavlovskii, 1956, *J. Exptl. Theoret. Phys. (USSR)* **31**, 427–438 [Engl. transl. “The self-consistent field equations in an atom,” *Sov. Phys. JETP* **4**, 328–336 (1957)].
- Lenard, A., and F. J. Dyson, 1968, “Stability of matter. II,” *J. Math. Phys.* **9**, 698–711.
- Lenz, W., 1932, “Über die Anwendbarkeit der statistischen Methode auf Ionengitter,” *Z. Phys.* **77**, 713–721.
- Lieb, E. H., 1976, “Bounds on the eigenvalues of the Laplace and Schrödinger operators,” *Bull. Am. Math. Soc.*, in press.
- Lieb, E. H., and J. L. Lebowitz, 1972, “The constitution of matter: existence of thermodynamics for systems composed of electrons and nuclei,” *Adv. Math.* **9**, 316–398. See also J. L. Lebowitz and E. H. Lieb, 1969, “Existence of thermodynamics for real matter with Coulomb forces,” *Phys. Rev. Lett.* **22**, 631–634.
- Lieb, E. H., and H. Narnhofer, 1975, “The thermodynamic limit for jellium,” *J. Stat. Phys.* **12**, 291–310; Erratum: *J. Stat. Phys.* **14**, No. 5 (1976).
- Lieb, E. H., and B. Simon, “On solutions to the Hartree–Fock problem for atoms and molecules,” *J. Chem. Phys.* **61**, 735–736. Also, a longer paper in preparation.
- Lieb, E. H., and B. Simon, 1977, “The Thomas–Fermi theory of atoms, molecules and solids,” *Adv. Math.*, in press. See also E. H. Lieb and B. Simon, 1973, “Thomas–Fermi theory revisited,” *Phys. Rev. Lett.* **33**, 681–683.
- Lieb, E. H., and W. E. Thirring, 1975, “A bound for the kinetic energy of fermions which proves the stability of matter,” *Phys. Rev. Lett.* **35**, 687–689; Errata: *Phys. Rev. Lett.* **35**, 1116. For more details on kinetic energy inequalities and their application, see also E. H. Lieb and W. E. Thirring, 1976, “Inequalities for the moments of the Eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities,” in *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*, edited by E. H. Lieb, B. Simon, and A. S. Wightman (Princeton University, Princeton).
- Rosen, G., 1971, “Minimum value for c in the Sobolev inequality $\|\phi\|_3 \leq c \|\nabla \phi\|^2$,” *SIAM J. Appl. Math.* **21**, 30–32.
- Schwinger, J., 1961, “On the bound states of a given potential,” *Proc. Nat. Acad. Sci. (U.S.)* **47**, 122–129.
- Scott, J. M. C., 1952, “The binding energy of the Thomas Fermi atom,” *Phil. Mag.* **43**, 859–867.
- Sheldon, J. W., 1955, “Use of the statistical field approximation in molecular physics,” *Phys. Rev.* **99**, 1291–1301.
- Slater, J. C., 1930, “The theory of complex spectra,” *Phys. Rev.* **34**, 1293–1322.
- Sobolev, S. L., 1938, *Mat. Sb.* **46**, 471. See also S. L. Sobolev, 1950, “Applications of functional analysis in mathematical physics,” *Leningrad; Am. Math. Soc. Transl. Monographs* **7** (1963).
- Sommerfeld, A., 1932, “Asymptotische Integration der Differentialgleichung des Thomas–Fermischen Atoms,” *Z. Phys.* **78**, 283–308.
- Teller, E., 1962, “On the stability of molecules in the Thomas–Fermi theory,” *Rev. Mod. Phys.* **34**, 627–631.
- Thomas, L. H., 1927, “The calculation of atomic fields,” *Proc. Camb. Phil. Soc.* **23**, 542–548.
- Von Weizsäcker, C. F., 1935, “Zur Theorie der Kernmassen,” *Z. Phys.* **96**, 431–458.

EXISTENCE OF THERMODYNAMICS FOR REAL MATTER WITH COULOMB FORCES

J. L. Lebowitz*

Belfer Graduate School of Science, Yeshiva University, New York, New York 10033

and

Elliott H. Lieb†

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 3 February 1969)

It is shown that a system made up of nuclei and electrons, the constituents of ordinary matter, has a well-defined statistical-mechanically computed free energy per unit volume in the thermodynamic (bulk) limit. This proves that statistical mechanics, as developed by Gibbs, really leads to a proper thermodynamics for macroscopic systems.

In this note we wish to report the solution to a classic problem lying at the foundations of statistical mechanics.

Ever since the daring hypothesis of Gibbs and others that the equilibrium properties of matter could be completely described in terms of a phase-space average, or partition function, $Z = \text{Tr} e^{-\beta H}$, it was realized that there were grave difficulties in justifying this assumption in terms of basic microscopic dynamics and that such delicate matters as the ergodic conjecture stood in the way. These questions have still not been satisfactorily resolved, but more recently still another problem about Z began to receive attention: Assuming the validity of the partition function, is it true that the resulting properties of matter will be extensive and otherwise the same as those postulated in the science of thermodynamics? In particular, does the thermodynamic, or bulk, limit exist for the free energy derived from the

partition function, and if so, does it have the appropriate convexity, i.e., stability properties?

To be precise, if N_j are an unbounded, increasing sequence of particle numbers, and Ω_j a sequence of reasonable domains (or boxes) of volume V_j such that $N_j/V_j \rightarrow \text{constant} = \rho$, does the free energy per unit volume

$$f_j = -kT(V_j)^{-1} \ln Z(\beta, N_j, \Omega_j) \quad (1)$$

approach a limit [called $f(\beta, \rho)$] as $j \rightarrow \infty$, and is this limit independent of the particular sequence and shape of the domains? If so, is f convex in the density ρ and concave in the temperature β^{-1} ? Convexity is the same as thermodynamic stability (non-negative compressibility and specific heat).

Various authors have evolved a technique for proving the above,^{1,2} but always with one severe drawback. It had to be assumed that the interparticle potentials were short range (in a manner to

be described precisely later), thereby excluding the Coulomb potential which is the true potential relevant for real matter. In this note we will indicate the lines along which a proof for Coulomb forces can be and has been constructed. The proof itself, which is quite long, will be given elsewhere.³ We will also list here some additional results for charged systems that go beyond the existence and convexity of the limiting free energy.

To begin with, a *sine qua non* for thermodynamics is the stability criterion on the N -body Hamiltonian $H = E_K + V$. It is that there exists a constant $B \neq 0$ such that for all N ,

$$V(r_1, \dots, r_N) > -BN \quad (\text{classical mechanics}), \quad (2)$$

$$E_0 > -BN \quad (\text{quantum mechanics}), \quad (3)$$

where E_0 is the ground-state energy in infinite space. (Classical stability implies quantum-mechanical stability, but not conversely.) Heuristically, stability insures against collapse. From the mathematical point of view, it provides a lower bound to f_j in (1). We wish to emphasize that stability of the Hamiltonian (H stability), while necessary, is insufficient for assuring the existence of thermodynamics. For example, it is trivial to prove H stability for charged particles all of one sign, and it is equally obvious that the thermodynamic limit does not exist in this case.

It is not too difficult to prove classical and thus also quantum-mechanical H stability for a wide variety of short-range potentials or for charged particles having a hard core.^{2,4} But real charged particles require quantum mechanics and the recent proof of H stability by Dyson and Lenard⁵ is as difficult as it is elegant. They show that stability will hold for any set of charges and masses provided that the negative particles and/or the positive ones are fermions.

The second requirement in the canonical proofs¹ is that the potential be tempered, which is to say that there exist a fixed r_0 and constants $C \geq 0$ and $\epsilon > 0$ such that if two groups of N_a and N_b particles are separated by a distance $r > r_0$, their interparticle energy is bounded by

$$V(N_a \oplus N_b) - V(N_a) - V(N_b) \leq C r^{-(3+\epsilon)} N_a N_b. \quad (4)$$

Tempering is roughly the antithesis of stability

because the requirements that the forces are not too repulsive at infinity insures against "explosion." Coulomb forces are obviously not tempered and for this reason the canonical proofs have to be altered. Our proof, however, is valid for a mixture of Coulomb and tempered potentials and this will always be understood in the theorems below. It is not altogether useless to include tempered potentials along with the true Coulomb potentials because one might wish to consider model systems in which ionized molecules are the elementary particles.

Prior to explaining how to overcome the lack of tempering we list the main theorems we are able to prove. These are true classically as well as quantum mechanically. But first three definitions are needed:

(D1) We consider s species of particles with charges e_i , particle numbers $N^{(i)}$, and densities $\rho^{(i)}$. In the following N and ρ are a shorthand notation for s -fold multiplets of numbers. The conditions for H stability (see above) are assumed to hold.

(D2) A neutral system is one for which $\sum_1^s N^{(i)} \times e_i = 0$, alternatively $\sum_1^s \rho^{(i)} e_i = 0$.

(D3) The ordinary s -species grand canonical partition function is

$$\sum_{N^{(s)}=0}^{\infty} \cdots \sum_{N^{(1)}=0}^{\infty} \prod_i^s z_i^{N^{(i)}} Z(N, \Omega). \quad (5)$$

The neutral grand canonical partition function is the same as (5) except that only neutral systems enter the sum.

The theorems are the following:

(T1) The canonical, thermodynamic limiting free energy per unit volume $f(\beta, \rho)$ exists for a neutral system and is independent of the shape of the domain for reasonable domains. Furthermore, $f(\beta, \rho^{(1)}, \rho^{(2)}, \dots)$ is concave in β^{-1} and jointly convex in the s variables $(\rho^{(1)}, \dots, \rho^{(s)})$.

(T2) The thermodynamic limiting microcanonical⁶ entropy per unit volume exists for a neutral system and is a concave function of the energy per unit volume. It is also independent of domain shape for reasonable shapes and it is equal to the entropy computed from the canonical free energy.

(T3) The thermodynamic limiting free energy per unit volume exists for both the ordinary and the neutral grand canonical ensembles and are independent of domain shape for reasonable domains. Moreover, they are equal to each other

and to the neutral canonical free energy per unit volume.

Theorem 3 states that systems which are not charge neutral make a vanishingly small contribution to the grand canonical free energy. While this is quite reasonable physically, it does raise an interesting point about nonuniform convergence because the ordinary and neutral partition functions are definitely not equal if we switch off the charge before passing to the thermodynamic limit, whereas they are equal if the limits are taken in the reverse order.

An interesting question is how much can charge neutrality be nonconserved before the free energy per unit volume deviates appreciably from its neutral value? The answer is in theorem 4.

(T4) Consider the canonical free energy with a surplus (i.e., imbalance) of charge Q and take the thermodynamic limit in either of three ways: (a) $QV^{-2/3} \rightarrow 0$; (b) $QV^{-2/3} \rightarrow \infty$; (c) $QV^{-2/3} \rightarrow \text{const.}$ In case (a) the limit is the same as for the neutral system while in case (b) the limit does not exist, i.e., $f \rightarrow \infty$. In case (c) the free energy approaches a limit equal to the neutral-system free energy plus the energy of a surface layer of charge Q as given by elementary electrostatics.

We turn now to a sketch of the method of proof and will restrict ourselves here to the neutral canonical ensemble. As usual, one first proves the existence of the limit for a standard sequence of domains. The limit for an arbitrary domain is then easily arrived at by packing that domain with the standard ones. The basic inequality that is needed is that if a domain Ω containing N particles is partitioned into D domains $\Omega_1, \Omega_2, \dots, \Omega_D$ containing $N_1, N_2, \dots, N_D$ particles, respectively, and if the interdomain interaction be neglected, then

$$Z(N, \Omega) \geq \prod_1^D Z(N_i, \Omega_i). \quad (6)$$

If Ω is partitioned into subdomains, as above, plus "corridors" of thickness $> r_0$ which are devoid of particles, one can use (4) to obtain a useful bound on the tempered part of the omitted interdomain interaction energy. We will refer to these energies as surface terms.

The normal choice¹ for the standard domains are cubes C_j containing N_j particles, with C_{j+1} being composed of eight copies of C_j together with corridors, and with $N_{j+1} = 8N_j$. Neglecting surface terms one would have from (6) and (1)

$$f_{j+1} \leq f_j. \quad (7)$$

Since f_j is bounded below by H stability, (7) implies the existence of a limit. To justify neglect of the surface terms one makes the corridors increase in thickness with increasing j ; although V_j^C , the corridor volume, approaches ∞ one makes $V_j^C/V_j \rightarrow 0$ in order that the limiting density not vanish. The positive ϵ of (4) allows one to accomplish these desiderata.

Obviously, such a strategy will fail with Coulomb forces, but fortunately there is another way to bound the interdomain energy. The essential point is that it is not necessary to bound this energy for all possible states of the systems in the subdomains; it is only necessary to bound the "average" interaction between domains, which is much easier. This is expressed mathematically by using the Peierls-Bogoliubov inequality⁷ to show that

$$Z(N, \Omega) \geq e^{-\beta U} \prod_1^D Z(N_i, \Omega_i), \quad (8)$$

where U is the average interdomain energy in an ensemble where each domain is independent. U consists of a Coulomb part, U_C , and a tempered part, U_t , which can be readily bounded.¹

We now make the observation, which is one of the crucial steps in our proof, that independently of charge symmetry U_C will vanish if the subdomains are spheres and are overall neutral. The rotation invariance of the Hamiltonian will produce a spherically symmetric charge distribution in each sphere and, as Newton⁸ observed, two such spheres would then interact as though their total charges (which are zero) were concentrated at their centers.

With this in mind we choose spheres for our standard domains. Sphere S_j will have radius $R_j = \rho^j$ with ρ an integer. The price we pay for using spheres instead of cubes is that a given one, S_k , cannot be packed arbitrarily full with spheres S_{k-1} only. We prove, however, that it can be packed arbitrarily closely (as $k \rightarrow \infty$) if we use all the previous spheres $S_{k-1}, S_{k-2}, \dots, S_0$. Indeed for the sequence of integers $n_1, n_2, \dots, n_j = (\rho-1)^{j-1} \rho^{2j}$ we can show that we can simultaneously pack n_j spheres S_{k-j} into S_k for $1 \leq j \leq k$. The fractional volume of S_k occupied by the S_{k-j} spheres is $\varphi_j = \rho^{-3j} n_j$, and from (8) we then have

$$f_k \leq \varphi_1 f_{k-1} + \varphi_2 f_{k-2} + \dots + \varphi_k f_0, \quad (9)$$

and

$$\sum_1^{\infty} \varphi_j = 1. \quad (10)$$

[Note that the inequality (6) is correct as it stands for pure Coulomb forces because U_C in (8) is identically zero. If short-range potentials are included there will also be surface terms, as in the cube construction, but these present only a technical complication that can be handled in the same manner as before.¹] While Eq. (9) is more complicated than (7), it is readily proven explicitly that f_k approaches a limit as $k \rightarrow \infty$. [Indeed, it follows from the theory of the renewal equation⁹ that (9) will have a limit if $\sum_1^\infty j\psi_j < \infty$.]

The possibility of packing spheres this way is provided by the following geometrical theorem which plays the key role in our analysis. We state it without proof, but we do so in d dimensions generally and use the following notation: σ_d = volume of a unit d -dimensional sphere = $\frac{4}{3}\pi$ in three dimensions and $\alpha_d = (2^d - 1)2d^{\frac{d-1}{2}}$.

(T5) Let $p \leq \alpha_d + 2^d \sigma_d^{-1}$ be a positive integer. For all positive integers j , define radii $r_j = p^{-j}$ and integers $n_j = (p-1)^j - 1 p^j (d-1)$. Then it is possible to place simultaneously $\bigcup_j (n_j \text{ spheres of radius } r_j)$ into a unit d -dimensional sphere so that none of them overlap.

The minimum value of p required by the theorem in three dimensions is 27.

Many of the ideas presented here had their genesis at the Symposium on Exact Results in Statistical Mechanics at Irvine, California, in 1968, and we should like to thank our colleagues for their encouragement and stimulation: M. E.

Fisher, R. Griffiths, O. Lanford, M. Mayer, D. Ruelle, and especially A. Lenard.

*Work supported by Air Force Office of Scientific Research, U. S. Air Force under Grant No. AFOSR 68-1416.

†Work supported by National Science Foundation Grant No. GP-9414.

¹These developments are clearly expounded in M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 377 (1964); D. Ruelle, Statistical Mechanics (W. A. Benjamin, Inc., New York, 1969). For a synopsis, see also J. L. Lebowitz, Ann. Rev. Phys. Chem. **19**, 389 (1968).

²R. B. Griffiths, Phys. Rev. **176**, 655 (1968), and footnote 6a in A. Lenard and F. J. Dyson [J. Math. Phys. **9**, 698 (1968)]; O. Penrose, in Statistical Mechanics, Foundations and Applications, edited by T. Bak (W. A. Benjamin, Inc., New York, 1967), p. 98.

³E. H. Lieb and J. L. Lebowitz, "The Constitution of Matter," to be published.

⁴L. Onsager, J. Phys. Chem. **43**, 189 (1939); M. E. Fisher and D. Ruelle, J. Math. Phys. **7**, 260 (1966).

⁵F. J. Dyson and A. Lenard, J. Math. Phys. **8**, 423 (1967); A. Lenard and F. J. Dyson, J. Math. Phys. **9**, 698 (1968); F. J. Dyson, J. Math. Phys. **8**, 1538 (1967).

⁶R. B. Griffiths, J. Math. Phys. **6**, 1447 (1965).

⁷K. Symanzik, J. Math. Phys. **6**, 1155 (1965).

⁸I. Newton, in Mathematical Principles, translated by A. Motte, revised by F. Cajori (University of California Press, Berkeley, Calif., 1934), Book 1, p. 193, propositions 71, 76.

⁹W. Feller, An Introduction to Probability Theory and Its Applications (J. Wiley & Sons, New York, 1957), 2nd ed. Vol. 1, p. 290.

Note: This paper is the announcement of the existence of the thermodynamic limit for particles interacting via Coulomb forces. The full version of this work appears in item 58 of the list of publications:

E.H. Lieb and J.L. Lebowitz, *The Constitution of Matter: Existence of Thermodynamics for Systems Composed of Electrons and Nuclei*, Adv. in Math. **9**, 316-398 (1972).

An abridged version (item 65 in the list of publications) appeared in the first and second editions of this *Selected* but is omitted in the third edition for space reasons. A different abridged version also appears in Section V of the paper, *The Stability of Matter*, Rev. Mod. Phys. **48**, 553-569 (1976), which is included as item VI.1 in this edition.

The Thermodynamic Limit for Jellium

Elliott H. Lieb^{1,2} and Heide Narnhofer^{3,4}

Received December 27, 1974

The thermodynamic limit of the free energy, energy, pressure, and entropy is established for a neutral system of charged particles interacting with a fixed, uniformly charged background (jellium).

KEY WORDS: Thermodynamic limit; jellium; charged particles; uniform background; neutral system; free energy density; quantum mechanics; equilibrium statistical mechanics.

1. INTRODUCTION

In 1938 Wigner⁽¹⁾ introduced a model for matter which is now called jellium. One supposes that the electrons in a solid provide a uniform, constant charge background in which the heavier nuclei move. The Hamiltonian for the system consisting of N particles with coordinates $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ in a three-dimensional domain Λ is

$$H = (2m)^{-1} \sum_{i=1}^N p_i^2 + e^2 U(\mathbf{X})$$

$$U(\mathbf{X}) = \sum_{i < j}^N |\mathbf{x}_i - \mathbf{x}_j|^{-1} - \rho \sum_{i=1}^N \varphi(\mathbf{x}_i) + \frac{1}{2} \rho^2 \int_{\Lambda} \varphi(\mathbf{x}) d\mathbf{x} \quad (1)$$

Work partially supported by the National Science Foundation, grant GP31674X.

¹ On leave from Departments of Mathematics and Physics, MIT, Cambridge, Massachusetts.

² Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey.

³ Bell Laboratories, Murray Hill, New Jersey.

⁴ On leave from Institute for Theoretical Physics, University of Vienna, Austria.

and where

$$\rho\varphi(\mathbf{x}) = \rho \int_{\Lambda} |\mathbf{x} - \mathbf{y}|^{-1} d\mathbf{y}$$

is the Coulomb potential produced by the background of charge density ρ .

Throughout the following we shall set $m = e^2 = 1$, and $\hbar = 1$ in the quantum case. Thus the Bohr radius is equal to unity and the energy unit, the Rydberg (Ry), is equal to one-half. The dimensionless length r_s is equal to $[3/(4\pi\rho)]^{1/3}$. Whenever the distinction is necessary, we shall assume $\rho > 0$ and that the particles are negative.

We shall show that for neutral systems, i.e., $\rho|\Lambda| = N$, the thermodynamic functions per unit volume (free energy, energy, entropy, pressure) exist as $\Lambda \rightarrow \infty$.

It is also possible to consider the one- and two-dimensional versions of this problem, where the Coulomb potential $|\mathbf{x}|^{-1}$ is replaced by $-|x|$ and $-\ln|x|$, respectively. In the one-dimensional, classical case, Baxter⁽²⁾ calculated the partition function exactly. For that case, Kunz⁽³⁾ showed that the one-particle distribution function exists and that it has crystalline ordering, i.e., the Wigner lattice exists for all temperatures. Brascamp and Lieb⁽⁴⁾ showed the same to be true in the quantum mechanical case for one-component fermions when β is large enough. Although we do not deal with the one-dimensional problem here, our methods would apply in that case. In two dimensions there are difficulties connected with the long-range nature of the $-\ln|x|$ potential, and we shall not discuss this here.

The problem of jellium is closely related to the same problem for real matter treated by Lebowitz and Lieb^{(5,6),5} and their methods will be employed here. The difficulty with jellium is that the background is held rigid by definition and one cannot freely constrain the particles to lie in balls without at the same time imparting an enormous electrostatic energy to the system. On the other hand, the fixed background considerably simplifies the H -stability question. (Cf. Dyson and Lenard.⁽⁸⁾) The connection between the jellium and the real matter problems is discussed by Narnhofer and Thirring.⁽⁹⁾

In Section 2 we use H -stability to get an upper bound on the partition function Z . The H -stability itself is proved in the appendix.

Section 3 deals with the classical case. We first treat a distinguished sequence of domains, which are balls, and then we treat general domains. The usual results are obtained, except that since the free energy is not a convex function of the density for jellium, the compressibility can be negative and the grand canonical ensemble is not equivalent to the canonical ensemble.

⁵ See also Penrose and Smith.⁽⁷⁾

As in Ref. 5, we show that a system with an excess charge $Q \sim |\Lambda|^{2/3}$ has an excess free energy $-(2\beta)^{-1}Q^2/C$, where C is the capacity of Λ .

In Section 4 we outline the proof when weakly tempered potentials are also present. Although the thermodynamic limit exists in this case, we lose continuity in ρ —at least by our methods. This is an open question. The inclusion of hard cores is also not covered by our method and this, too, is an open question.

In Section 5 we explain the additional techniques needed for the quantum case. An open question here is to show the equivalence of different boundary conditions; we use Dirichlet conditions. A related problem is to show that the particle density and the electrostatic potential stay suitably bounded as $N \rightarrow \infty$.

2. H -STABILITY

The condition of H -stability is that the Hamiltonian is bounded below by a constant times N . It is sufficient to require that the potential energy alone has this property, since the kinetic energy operator is positive. For real matter one is obliged to consider the total Hamiltonian because the interaction energy of a positive and a negative particle has no lower bound. The proof of H -stability in this latter case is very difficult and was given by Dyson and Lenard^(8,10) and recently a new proof was given by Federbush.⁽¹¹⁾ It is essential here that the electrons be fermions, thereby excluding classical particles.

For jellium, on the other hand, one can easily find a lower bound on U , by using an idea due to Onsager.⁽¹²⁾ This is given in the appendix. A different proof and a different bound are also given in Ref. 10. Our bound is

$$U > -0.9N/r_s \quad (2)$$

and we emphasize that this result holds for all N and all domains, connected or not, and requires only that the background have charge density $(3/4\pi)r_s^{-3}$ or zero everywhere. This lower bound is surprisingly accurate. In Ref. 13 a numerical evaluation for the body-centered cubic lattice of particles in a uniform background gives

$$U_{\min} \leq -0.896N/r_s \quad (3)$$

when the system is neutral.

The significance of the lower bound, and the only place it will be used here, is to establish an upper bound for the partition function Z , i.e.,

$$Z \leq Z_{\text{ideal}} e^{zN} \quad (4)$$

where ξ is some constant and Z_{ideal} is the partition function of ideal, non-interacting particles. Thus, defining

$$g = V^{-1} \ln Z \quad (5)$$

for a domain of volume V , one has that g is bounded above.

3. CLASSICAL PARTICLES WITH PURELY COULOMB FORCES

3.1. Canonical Ensemble (Spherical Domains)

Fix the density ρ . Let $\{B_k\}_{k=0}^\infty$ be a sequence of balls of radii $R_k = R_0(1+p)^k$, where $p = 26$ and the volume of $B_0 \equiv |B_0|$ is ρ^{-1} . Let $N_k = (1+p)^{3k}$ be the number of particles in B_k , whence $\rho_k = N_k/|B_k| = \rho$. Let $n_j = p^{j-1}(1+p)^{2j}$. According to Ref. 5, Section III, one can pack B_K with $\cup_{j=0}^{K-1} (n_{K-j} \text{ balls } B_j)$ so that they do not overlap, and

$$\lim_{K \rightarrow \infty} |B_K|^{-1} \sum_{j=0}^{K-1} n_{K-j} |B_j| = 1 \quad (6)$$

The part of B_K not covered by the above packing will be called D_K .

At this point the principal difference between the proof for the jellium model and the proof for a system of positive and negative particles appears. In the latter, the N_K particles are constrained to be in the balls B_j , $j < K$, and the domain D_K is left empty. For jellium this cannot be done because the domain D_K would then not be neutral and the electrostatic energy of the system would be too large. Even though $|D_K|/|B_K| \rightarrow 0$ as $K \rightarrow \infty$, N_K^{-1} (the electrostatic energy of D_K) would go to infinity.

We proceed as follows: Let Z_k , $k = 0, 1, 2, \dots$, be the configurational partition function of the ball B_k with N_k particles and with a uniform background of density ρ :

$$Z_k = (N_k!)^{-1} \int_{(B_k)^{N_k}} \exp[-\beta U(\mathbf{x}_1, \dots, \mathbf{x}_{N_k})] d\mathbf{x}_1 \dots d\mathbf{x}_{N_k} \quad (7)$$

Let Z_K^D be the configurational partition function of D_K with M_K particles, where

$$M_K = N_K - \sum_{j=0}^{K-1} n_{K-j} N_j = N_K p^K (1+p)^{-K} \quad (8)$$

D_K is understood to have a uniform background of density ρ . Clearly, $\rho D_K = M_K$ and $M_K/N_K \rightarrow 0$ exponentially fast.

The fundamental inequality, to be found in Ref. 5, Section IIE, is that

$$\ln Z_K \geq \sum_{j=0}^{K-1} n_{K-j} \ln Z_j + \ln Z_K^D \quad (9)$$

This inequality exploits Newton's electrostatic theorem and the fact that all the subdomains, except D_K , are both spherical and neutral; therefore the average interdomain interaction is zero.

The next step is to estimate Z_K^D . Using Jensen's inequality,

$$\ln Z_K^D \geq M_K \ln |D_K| - \ln(M_K!) - \beta \langle U \rangle_{D_K}$$

where

$$\begin{aligned} \langle U \rangle_{D_K} &= \frac{1}{2} M_K (M_K - 1) |D_K|^{-2} \int \int_{D_K} |\mathbf{x} - \mathbf{y}|^{-1} d\mathbf{x} d\mathbf{y} \\ &\quad + \frac{1}{2} \rho^2 \int \int_{D_K} |\mathbf{x} - \mathbf{y}|^{-1} d\mathbf{x} d\mathbf{y} \\ &\quad - \rho M_K |D_K|^{-1} \int \int_{D_K} |\mathbf{x} - \mathbf{y}|^{-1} d\mathbf{x} d\mathbf{y} \end{aligned} \quad (10)$$

Since $M_K = \rho \int_{D_K} d\mathbf{x}$,

$$\langle U \rangle_{D_K} = -\frac{1}{2} \rho |D_K|^{-1} \int \int_{D_K} |\mathbf{x} - \mathbf{y}|^{-1} d\mathbf{x} d\mathbf{y} < 0 \quad (11)$$

Thus, defining

$$g_K = |B_K|^{-1} \ln Z_K \quad (12)$$

and

$$\gamma = p(1 + p)^{-1} < 1 \quad (13)$$

we have, for large K ,

$$g_K \geq p^{-1} \sum_{j=0}^{K-1} \gamma^{K-j} g_j + \gamma^K \rho (1 - \ln \rho) \quad (14)$$

where Stirling's formula for $M_K!$ has been used. As shown in Ref. 5, Section IVD, this inequality implies that g_K has a limit, $g(\beta, \rho)$, for this special, ρ -dependent sequence of domains B_K .

3.2. Canonical Ensemble (General Domains)

Let ρ be fixed. We take a regular sequence of domains $\{\Lambda_j\}_{j=1}^\infty$ tending to infinity which satisfies conditions A (Van Hove limit) and B (ball condition) given in Ref. 5, Section V, and which *also* satisfies the condition that $\rho|\Lambda_j| = j$. To get a lower bound on $Z(\Lambda_j)$, we pack Λ_j with balls B_k of the standard sequence appropriate to ρ given above and distribute the j particles with constant density in the balls and in D_j , which is the complement of the B_k in Λ_j . As above, we have

$$|\Lambda_j| g_j \equiv \ln Z(\Lambda_j) \geq \sum_k m_{jk} \ln Z_k + \ln Z_j^D \quad (15)$$

where m_{jk} is the number of balls B_k in the packing of Λ_j . Met M_j be the number of particles in D_j , i.e.,

$$M_j = j - \sum_k m_{jk}(1 + p)^{3k}$$

Then, as before,

$$\ln Z_j^p \geq M_j \ln |D_j| - \ln(M_j!) - \beta \langle U \rangle_{D_j} \quad (16)$$

and $\langle U \rangle_{D_j} \leq 0$. Following the proof in Ref. 5, Section V,

$$\liminf_{j \rightarrow \infty} g_j \geq g(\beta, \rho) \quad (17)$$

where $g(\beta, \rho)$ is the limit for the standard balls.

An upper bound to $Z(\Lambda_j)$ can be found by embedding Λ_j in a minimum standard ball $B_{K(j)}$ and packing $B_{K(j)} \setminus \Lambda_j$ with balls B_k . Let D_j be as before, i.e., $B_{K(j)} \setminus (\Lambda_j \cup B_k)$ and $M_j = \rho |D_j|$. Then

$$\begin{aligned} \ln Z_{K(j)} &\geq \ln Z(\Lambda_j) + \sum m'_{jk} \ln Z_k + M_j \ln |D_j| - \ln(M_j!) \\ &\quad - \beta \langle U(D_j, D_j) \rangle - \beta \langle U(D_j, \Lambda_j) \rangle \end{aligned} \quad (18)$$

In the last four terms we use Jensen's inequality for the integration over the coordinates of the particles in D_j : $\langle U(D_j, D_j) \rangle$ is the average Coulomb energy in D_j in an ensemble in which the particles are free; $\langle U(D_j, \Lambda_j) \rangle$ is the average interdomain interaction between D_j and Λ_j when the particles in D_j are free and the particles in Λ_j are fully interacting. The last term is zero because the average total charge distribution in D_j is zero. The term $\langle U(D_j, D_j) \rangle$ is negative as before. Thus we can use the argument of Ref. 5, Section V, to conclude that

$$\limsup_{j \rightarrow \infty} g_j \leq g(\beta, \rho) \quad (19)$$

The result of these inequalities is that for any regular sequence of domains $\{\Lambda_j\}$ and particle numbers $N_j = j$ such that $N_j = \rho |\Lambda_j|$,

$$\lim_{j \rightarrow \infty} g_j = g(\beta, \rho) \quad (20)$$

While this establishes the existence and shape independence of the thermodynamic limit for each fixed ρ , we do not yet know anything about the dependence of $g(\beta, \rho)$ on ρ or whether the limit is uniform in ρ . We next discuss how such a relationship can be obtained.

3.3. Scaling Relations

Let $\{\Lambda_j\}_{j=1}^\infty$ be a regular sequence of domains for a given ρ , i.e., $\rho |\Lambda_j| = j$. Let $\eta > 0$ be fixed and define the following:

$$\rho' = \rho \eta^3, \quad \beta' = \beta \eta^{-1}, \quad \Lambda'_j = \eta^{-1} \Lambda_j = \{\eta^{-1} \mathbf{x} | \mathbf{x} \in \Lambda_j\} \quad (21)$$

Thus $\rho'|\Lambda_j'| = j$.

If one considers the integral defining $Z(\beta, \rho; \Lambda_j)$ and changes integration variables $\mathbf{x}$ to $\mathbf{y} = \eta^{-1}\mathbf{x}$, then one derives

$$|\Lambda_j|g(\beta, \rho; \Lambda_j) = |\Lambda_j'|g(\beta\eta^{-1}, \rho\eta^3; \Lambda_j') + 3j \ln \eta \quad (22)$$

Since the thermodynamic limit is independent of the sequence of domains, one has that

$$g(\beta, \rho) = \eta^{-3}g(\beta\eta^{-1}, \rho\eta^3) + 3\rho \ln \eta \quad (23)$$

for all $\eta > 0$. Now let $\eta = \rho^{-1/3}$, whence

$$g(\beta, \rho) = \rho g(\beta\rho^{1/3}, 1) - \rho \ln \rho = \rho \bar{g}(\beta\rho^{1/3}) + \rho(1 - \ln \rho) \quad (24)$$

where $\bar{g}(\cdot) \equiv g(\cdot, 1) - 1$.

From the basic definition of $g(\beta, \rho; \Lambda_j)$ one has that these functions, and hence their limits also, are convex functions of β . Therefore the function $t \rightarrow \bar{g}(t)$ is convex in t .

For finite j , let $\{\Lambda_j'\}$ be a regular sequence of domains with $|\Lambda_j'| = j$, and define

$$\bar{g}_j(\beta) = g(\beta, 1; \Lambda_j') - 1 \quad (25)$$

Then

$$\lim_{j \rightarrow \infty} \bar{g}_j(\beta) = \bar{g}(\beta) \quad (26)$$

3.4. Properties of the Thermodynamic Limit

3.4.1. Uniformity of the Limit. Since $\bar{g}(t)$ is bounded on finite t intervals, its convexity implies that it is continuous. Furthermore, each $\bar{g}_j(\cdot)$ has the same properties from (25). Thus the sequence of functions $\bar{g}_j(\beta\rho^{1/3})$ is continuous in ρ and has a continuous limit $\bar{g}(\beta\rho^{1/3})$ and the limit is essentially monotone as one sees from (14). Hence Dini's theorem tells us that the limit is uniform on compact ρ intervals.

3.4.2. Pressure and Compressibility. For a normal thermodynamic system, $g(\beta, \rho)$ is *concave* in ρ . This implies positive compressibility and, since the pressure is zero at zero density, it implies positive pressure. For jellium this is unfortunately not true. Using (24), and assuming differentiability, we obtain

$$\beta P/\rho = 1 - \frac{1}{3}\beta\rho^{1/3}\dot{\bar{g}}(\beta\rho^{1/3}) \quad (27)$$

where the dot denotes derivative, and

$$\beta\kappa^{-1} = \beta \frac{dP}{d\rho} = 1 - \frac{4}{9}\beta\rho^{1/3}\dot{\bar{g}}(\beta\rho^{1/3}) - \frac{\beta^2}{9}\rho^{2/3}\ddot{\bar{g}}(\beta\rho^{1/3}) \quad (28)$$

Note that $\ddot{g} > 0$. This implies that for $t > 0$, $\dot{g}(t) \geq \dot{g}(0) = 0$. These formulas show that P and κ can have either sign. In fact, for fixed β , they are both negative for sufficiently high density since, from (3), one sees that the potential energy will go as $\rho^{4/3}$ for large ρ , i.e., $g(t) \sim t$ for large t . Since $\dot{g}(t)$ is monotone, $t\dot{g}(t)$ is also monotone. This implies that there is always exactly one value, $(\beta\rho^{1/3})_c$, of $\beta\rho^{1/3}$ at which the pressure is zero. Without any constraint on the volume, classical jellium would collapse to a density $\rho^{1/3} = (\beta\rho^{1/3})_c\beta^{-1}$. This fact is not unrelated to the absence of H -stability for real matter without Fermi statistics.

3.5. Systems That Are Not Neutral

We wish to consider a sequence of systems with fixed background density ρ , but where $N \neq \rho|\Lambda|$. Define $Q_j \equiv -N_j + \rho|\Lambda_j|$ to be the net charge in Λ_j , and consider a sequence of domains Λ_j of *fixed shape* of capacitance $C_j = c|\Lambda_j|^{1/3}$. If $Q_j|\Lambda_j|^{-2/3} \rightarrow \sigma$, the result to be proved is that

$$g_j(\beta, \rho) \rightarrow g(\beta, \rho) - \frac{1}{2}\sigma^2c \quad (29)$$

Note that σ can have either sign. If $|\sigma| = \infty$, then $g_j(\beta, \rho) \rightarrow -\infty$. This last statement is easily proved by noting that $|\Lambda_j|^{-1} \min\{U(\mathbf{x})|\mathbf{x}_i \in \Lambda_j\} \rightarrow +\infty$ when $|Q_j||\Lambda_j|^{-2/3} \rightarrow +\infty$.

In order to simplify matters we shall prove the theorem only for balls, in which case $c = (4\pi/3)^{-1/3}$.

Let B be a ball of radius R and let B' be a concentric ball of radius $R' > R$. Note that a uniform charge density τ placed in $\Sigma \equiv B' \setminus B$ produces a constant potential $\tau\Phi(\Sigma)$ inside B . This same charge density in Σ has a self-energy $\tau^2S(\Sigma)$. If $R \rightarrow \infty$ and $R'/R \rightarrow 1$, then

$$\Phi(\Sigma)|\Sigma|/S(\Sigma) \rightarrow 2 \quad (30)$$

Let $Z(N, B')$ be the partition function for N particles in B' with background density ρ . A lower bound to $Z(N, B')$ can be obtained as follows:

1. Restrict the configurations to N_1 particles in B and $N_2 = N - N_1$ particles in Σ .

2. Let $U_1(\mathbf{X}_1)$ [resp. $U_2(\mathbf{X}_2)$] be the potential energy of the particles and background in B [resp. Σ] and let $U_{12}(\mathbf{X}_1, \mathbf{X}_2)$ be the interdomain energy, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are the particle coordinates. Then

$$\begin{aligned} Z(N, B') &\geq (N_1! N_2!)^{-1} \int_{B^{N_1}} \exp[-\beta U_1(\mathbf{X}_1)] \\ &\quad \times \int_{\Sigma^{N_2}} \exp\{-\beta[U_2(\mathbf{X}_2) + U_{12}(\mathbf{X}_1, \mathbf{X}_2)]\} \end{aligned} \quad (31)$$

3. Use Jensen's inequality on the second integral together with the aforementioned constancy of the potential $\Phi(\Sigma)$. Thus,

$$\begin{aligned} \ln Z(N, B') &\geq \ln Z(N_1, B) + \ln\{|\Sigma|^{N_2/N_2!}\} \\ &\quad - \beta S(\Sigma) \left[\frac{1}{2} \rho^2 + \binom{N_2}{2} |\Sigma|^{-2} - N_2 \rho |\Sigma|^{-1} \right] \\ &\quad - \beta \Phi(\Sigma) [\rho - N_2 |\Sigma|^{-1}] [\rho |B'| - N_1] \end{aligned} \quad (32)$$

Now we consider a sequence of balls B_j of radii R_j with background density ρ and particle numbers $N_j = j$, $j = 1, 2, \dots$. For $Q_j \equiv -j + \rho |B_j|$ negative we first use (32) with $N = j$, $B' = B_j$, $R = R_j - 1$, and $N_1 = \rho |B|$. Then we use (32) with $N = N_1 = j$, $B = B_j$, and $|B'| = j/\rho$. When $Q_j > 0$, we first use (32) with $N = N_1 = j$, $B' = B_j$, and $|B| = j/\rho$. Then we use (32) with $N_1 = j$, $B = B_j$, $R' = R_j + 1$, and $N = |B'| \rho$. Using the fact that $Q_j |B_j|^{-2/3} \rightarrow \sigma$ and (30), we obtain the desired result (29).

3.6. Microcanonical Ensemble

The existence of the thermodynamic limit for the microcanonical ensemble can be demonstrated using the methods of Ref. 5, Section VIII. There, the energy as a function of entropy was given for the quantum case. The corresponding classical equation is as follows: Let $\Gamma(N, \Lambda) = (\Lambda \times \mathbf{R}^3)^N$ be the phase space (including momentum). For σ real, let

$$\Delta(\sigma, N, \Lambda) = \{A \subset \Gamma(N, \Lambda) | \mu(A) = e^{\sigma |\Lambda|}\} \quad (33)$$

where μ is Lebesgue measure. Let

$$\epsilon(A, N, \Lambda) = |\Lambda|^{-1} \int_A H(\mathbf{X}, \mathbf{P}) e^{-\sigma |\Lambda|} \quad (34)$$

where

$$H(\mathbf{X}, \mathbf{P}) = U(\mathbf{X}) + \sum \mathbf{p}_i^2 / 2m$$

Then we define

$$\epsilon(\sigma, N, \Lambda) = \inf\{\epsilon(A, N, \Lambda) | A \in \Delta(\sigma, N, \Lambda)\} \quad (35)$$

to be the energy per unit volume as a function of the entropy per unit volume, σ .

Obviously, when Λ_1 and Λ_2 are disjoint,

$$\Delta(\sigma_1 + \sigma_2, N_1 + N_2, \Lambda_1 \cup \Lambda_2) \supset \Delta(\sigma_1, N_1, \Lambda_1) \times \Delta(\sigma_2, N_2, \Lambda_2) \quad (36)$$

Hence

$$\begin{aligned} &|\Lambda| \epsilon(\sigma_1 + \sigma_2, N_1 + N_2, \Lambda_1 \cup \Lambda_2) \\ &\leq |\Lambda_1| \epsilon(\sigma_1, N_1, \Lambda_1) + |\Lambda_2| \epsilon(\sigma_2, N_2, \Lambda_2) + \langle U(\Lambda_1, \Lambda_2) \rangle \end{aligned} \quad (37)$$

where $\langle U \rangle$ is the average over A_1 and A_2 of the interaction energy between Λ_1 and Λ_2 . [This may require passing to a subsequence in (35) for Λ_1 and Λ_2 .]

Now we are in the same position as in (16); the existence and appropriate convexity properties of $\epsilon(\sigma)$ and $\sigma(\epsilon)$ follow. See Ref. 5, Section VIII for details.

The one essential difference from the systems studied in Ref. 5, Section VIII is that for jellium we do not obtain convexity of ϵ as a function of ρ , but only as a function of σ . This lack does not alter the equivalence of the canonical and microcanonical ensembles.

3.7. The Grand Canonical Ensemble

If one considers the grand canonical ensemble (GCE) for fixed Λ , fixed background density ρ , and fixed chemical potential μ , then the GCE partition function Ξ will exist. From the results of Section 3.5 the thermodynamic limit of $\pi = |\Lambda|^{-1} \ln \Xi$ will exist and $\pi = \rho\mu + g(\rho)$ as in Theorem 7.1 of Ref. 5, Section VII. If, on the other hand, one defines Ξ for neutral jellium by requiring that $\rho = N/|\Lambda|$ for each N , then Ξ will diverge, even for finite Λ . This is a consequence of (3) that $g(N, \Lambda) \sim N^{4/3}$ for large N . In the quantum case with fermions, this divergence will not occur since the kinetic energy is proportional to $N^{5/3}$. Although the thermodynamic limit of π for neutral jellium would then exist, it would not be equivalent to the canonical partition ensemble because of the lack of convexity of the free energy in ρ .

4. ADDITIONAL POTENTIALS

As was shown in Ref. 5, additional short-range forces among the particles can be included without any conceptual difficulty, but with a great deal of technical difficulty, provided they are tempered and provided that these forces are integrable. This means that hard cores are excluded. We do not say that the thermodynamic limit does not exist when hard cores are present—it probably does—but only that our method is not adequate. The difficulty arises in (11), where $\ln Z_K^D$ is estimated by Jensen's inequality in terms of $\langle U \rangle$. We made this estimate in order to show that the energy of the particles in D_K was not too large. If some other method could be found to show this, then perhaps hard cores could be included, but in our estimate, $\langle U \rangle = +\infty$ and $\ln Z_K^D \geq -\infty$ when hard cores are present.

There is another serious difficulty when additional potentials, even nice ones, are present. The scaling relation of Section 3.3 does not hold, and hence the continuity with respect to ρ that was used in Section 3.4.1 cannot be established that way.

5. QUANTUM MECHANICAL PARTICLES

We first remark that it is immaterial for our purposes whether the particles are bosons or fermions. In contrast to the situation for real particles, H -stability (2) holds in the classical sense and therefore Fermi statistics is not required. We shall construct the proof for fermions and it will obviously be valid for bosons as well. Dirichlet boundary conditions will be employed, i.e., $\psi = 0$ on the boundary of Λ .

5.1. Canonical Ensemble (Spherical Domains)

By well-known arguments (see Ref. 5, Section II), a lower bound to Z_K can be obtained by constraining the particles to lie in various subdomains. In this way we arrive at precisely the same inequality (9) as for the classical case. The problem is to show that $\ln Z_K^D$ is not too small. For this purpose it would be sufficient to find one wave function ψ for the M_K particles in D_K such that $\langle H_K \rangle \equiv \langle \psi, H_K \psi \rangle < (\text{positive constant})M_K$, where H_K is the total Hamiltonian of the M_K particles in D_K , including the background self-energy. Then, by the Peierls–Bogoliubov inequality,

$$\ln Z_K^D \geq -\beta \langle H_K \rangle \quad (38)$$

A natural suggestion would be to take a determinantal wave function that vanishes on the boundary of D_K , but this will not work for the reason that the single-particle density will not be a constant and consequently the estimate $\langle U \rangle_{D_K} < 0$ [Eq. (11)] will not hold. On the other hand, suppose one could find M_K points $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_{M_K}\}$ in D_K such that:

- (a) $U(\mathbf{Y}) < (\text{positive constant})M_K$.
- (b) $|\mathbf{y}_i - \mathbf{y}_j| > 2h$ for some fixed $h > 0$.
- (c) The distance of $\mathbf{y}_i$ to the boundary of D_K is $> h$ for all i .

Then one could construct a product wave function in which the single-particle wave functions have support in balls of radius h centered at the $\mathbf{y}_i$ and which are spherically symmetric about the $\mathbf{y}_i$. The kinetic energy would be proportional to h^{-2} . Due to the peculiar shape of D_K , we are unable to find such a $\mathbf{Y}$. What Eq. (11) shows is that there certainly exists a $\mathbf{Y}$ satisfying condition (a) but we do not know if it satisfies (b) and (c).

It is in fact possible to find a $\mathbf{Y}$ such that condition (a) is satisfied and condition (b) is *effectively* satisfied. To do this, define

$$U'(\mathbf{Y}) = U(\mathbf{Y}) + \sum_{i < j}^{M_K} L(\mathbf{y}_i - \mathbf{y}_j) \quad (39)$$

where $U(\mathbf{Y})$ is the Coulomb potential as before and

$$\begin{aligned} L(\mathbf{y}) &= 2(\pi/|\mathbf{y}|)^2 & \text{for } |\mathbf{y}| \leq 1 \\ &= 0 & \text{for } |\mathbf{y}| > 1 \end{aligned} \quad (40)$$

Using (11), we obtain

$$\begin{aligned}\langle U' \rangle_{D_K} &= \langle U \rangle_{D_K} + \binom{M_K}{2} |D_K|^{-2} \int_{D_K} \int L(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &\leq \binom{M_K}{2} |D_K|^{-1} 8\pi^3 \leq 4\pi^3 M_K \rho\end{aligned}\quad (41)$$

Therefore there exists a $\mathbf{Y}$ such that

$$U'(\mathbf{Y}) \leq 4\pi^3 M_K \rho \quad (42)$$

Let $d_i = \frac{1}{2} \min\{1, \min_{j \neq i} |\mathbf{y}_i - \mathbf{y}_j|\}$. Construct a product trial function ψ using single-particle functions $\{\varphi_i\}_{i=1}^{M_K}$ centered at $\mathbf{y}_i$ and having support in a ball of radius d_i of the form

$$\varphi_i(\mathbf{x}) = (2\pi d_i)^{-1/2} |\mathbf{x} - \mathbf{y}_i|^{-1} \sin[\pi |\mathbf{x} - \mathbf{y}_i|/d_i] \quad (43)$$

The kinetic energy of φ_i is $(\pi/d_i)^2/2$. The potential energy of ψ can be evaluated as follows: The particle-particle energy is the same as if the particles were located at $\mathbf{y}_i$, by Newton's theorem. The interaction of a smeared-out particle with the background is changed by the amount

$$\begin{aligned}\rho \int_{|\mathbf{x}| < d_i} d\mathbf{x} \int_{|\mathbf{y}| < d_i} d\mathbf{y} |\mathbf{x} - \mathbf{y}|^{-1} [\varphi_i(\mathbf{x} + \mathbf{y})^2 - \delta(\mathbf{x})] \\ = \xi \rho d_i^2 \leq \xi \rho\end{aligned}\quad (44)$$

where ξ is a constant, assuming that the ball of radius d_i lies entirely in D_K . Thus the total energy $\langle \psi, H_K \psi \rangle$ is less than

$$U'(Y) + \xi \rho M_K + 2\pi^2 M_K \leq (2\pi^2) M_K [1 + 2\pi \rho + \xi \rho (2\pi^2)^{-1}] \quad (45)$$

This result is exactly what conditions (a) and (b) would give.

Condition (c) is more difficult, for it requires that the coordinates in $\mathbf{Y}$ are not too close to the boundary. If one tries to introduce

$$U''(\mathbf{Y}) = U'(\mathbf{Y}) + \sum_i d(\mathbf{y}_i)^{-2}$$

where $d(\mathbf{y}_i)$ is the distance to the boundary, one will find that $\langle U'' \rangle_{D_K} = \infty$ since $d(\mathbf{x})^{-2}$ is not integrable.

Since we are unable to deal with this problem directly, we shall modify our basic construction for the ball packing in such a way that the balls have a minimum spacing of some length independent of K .

Let $\{B_k\}_{k=0}^\infty$ be a sequence of balls of radii $R_k = R_0'(1+p)^k(1 - \frac{1}{2}\theta^k)$ with $\theta = (1+p)^{-1}$, $p = 26$, and R_0' chosen so that $\rho|B_0| = 1$. Let $N_k = \rho|B_k|$, whence N_k is an integer. As shown in Ref. 5, Section IV, it is possible to pack B_K with n_{K-j} balls B_j so that the distance of every ball to the boundary is not less than $4h$ and the distance between balls is not less than $8h$, where

The Thermodynamic Limit for Jellium

303

$h = R_0'(1 - \theta)/8$. As in Section 3.1, the part of B_K not covered by the packing will be called D_K .

Let us label the individual balls in the packing of B_K by a superscript i , namely B^i , and let R^i be its radius. Around B^i we construct two concentric, spherical shells S^i and T^i of radii $(R^i, R^i + h)$ and $(R^i + h, R^i + 2h)$, respectively. Inside B_K we also construct two concentric, spherical shells S_K and T_K of radii $(R_K - h, R_K)$ and $(R_K - 2h, R_K - h)$, respectively. All these shells are disjoint and lie in D_K and we denote by $D_K' \subset D_K$ the complement of the shells in D_K , and define $D_K'' = D_K' \cup T^i \cup T_K$.

We wish to find a $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_{M_K}\}$ with $\mathbf{y}_i \in D_K''$, and a corresponding product wave function ψ such that $\langle H_K \rangle$ is not too large. To this end, let

$$\begin{aligned} f(\mathbf{y}) &= 1, & \mathbf{y} \in D_K' \\ &= 0, & \mathbf{y} \notin D_K'' \\ &= f^i, & \mathbf{y} \in T^i \\ &= f_K, & \mathbf{y} \in T_K \end{aligned} \quad (46)$$

where

$$\begin{aligned} f^i &= 1 + [(R^i + h)^3 - (R^i)^3][(R^i + 2h)^3 - (R^i + h)^3]^{-1} < 2 \\ f_K &= 1 + [(R_K - h)^3 - (R_K - 2h)^3]^{-1}[R_K^3 - (R_K - h)^3] < 3 \end{aligned} \quad (47)$$

whence

$$\int_{T^i} (f - 1) = \int_{S^i} 1$$

and similarly for T_K , S_K , and $\int f = |D_K|$.

Let

$$F(\mathbf{Y}) = \prod_{i=1}^{M_K} f(\mathbf{y}_i)$$

and let

$$\langle U' \rangle_F = \int F(\mathbf{Y}) U'(\mathbf{Y}) / \int F(\mathbf{Y}) \quad (48)$$

The part involving L is [using (41)]

$$\left(\frac{M_K}{2} \right) |D_K|^{-2} \iint f(\mathbf{x}) f(\mathbf{y}) L(\mathbf{x} - \mathbf{y}) \leq 9 \cdot 4\pi^3 M_K \rho$$

since $f(\mathbf{x}) \leq 3$. The part involving $U(\mathbf{Y})$ is

$$\begin{aligned} &\frac{1}{2} \rho^2 \int_{D_K} d\mathbf{x} \int_{D_K} d\mathbf{y} |\mathbf{x} - \mathbf{y}|^{-1} [1 - f(\mathbf{x})][1 - f(\mathbf{y})] \\ &- \frac{1}{2} \rho |D_K|^{-1} \int_{D_K} d\mathbf{x} \int_{D_K} d\mathbf{y} |\mathbf{x} - \mathbf{y}|^{-1} f(\mathbf{x}) f(\mathbf{y}) \end{aligned} \quad (49)$$

The second term is negative. The first is the Coulomb energy of double shells, each pair of which is neutral and spherically symmetric. By Newton's theorem, this is just the sum of the self-energies of each pair. Let E_k be the self-energy of the two shells S and T surrounding a ball B_k in the packing and let W_K be the self-energy of the shells S_K and T_K . Then

$$\langle U' \rangle_F \leq 36\pi^3 M_K \rho + \sum_{j=0}^{K-1} n_{K-j} E_j + W_K \leq \text{const} \times M_K \quad (50)$$

The latter inequality comes from an elementary calculation of E_k and W_K .

The conclusion is that there exists a $\mathbf{Y}$ with $\mathbf{y}_i \in D_K''$ such that $U'(\mathbf{Y})$ is bounded by a constant times $|D_K|$. Now we construct a trial function ψ as before with φ_i given by (43) except that

$$d_i = \frac{1}{2} \min\{2h, 1, \min_{j \neq i} |\mathbf{y}_i - \mathbf{y}_j|\}$$

Then

$$\begin{aligned} |B_K|^{-1} \ln Z_K^D &\geq -\beta |B_K|^{-1} \langle \psi, H_K \psi \rangle \\ &\geq -\beta \times \text{const} \times |D_K| |B_K|^{-1} \end{aligned} \quad (51)$$

and this goes to zero as $K \rightarrow \infty$ like γ^K [Eq. (13)]. Thus the thermodynamic limit is established as in the classical case.

5.2. Canonical Ensemble (General Domains)

Let ρ be fixed. We take a regular sequence of domains $\{\Lambda_j\}_{j=1}^\infty$ tending to infinity which satisfies conditions A and B of Ref. 5, Section V. Also, $|\Lambda_j| \rho = j$. In addition, we require some conditions on the sequence which are not required in the classical case. These are the following:

- (i) Let $h > 0$ and let Λ_j^h and Λ_j^{2h} be the domains

$$\begin{aligned} \Lambda_j^h &= \{\mathbf{x} \in R^3 | \mathbf{x} \notin \Lambda_j, d(\mathbf{x}; \partial \Lambda_j) \leq h\} \\ \Lambda_j^{2h} &= \{\mathbf{x} \in R^3 | \mathbf{x} \notin \Lambda_j^h \cup \Lambda_j, d(\mathbf{x}; \partial \Lambda_j^h) \leq h\} \end{aligned} \quad (52)$$

where $d(\cdot; \cdot)$ is the Euclidean distance. We require that $|\Lambda_j^h|/|\Lambda_j^{2h}|$ be bounded in j for each fixed h .

- (ii) Consider the charge density

$$\begin{aligned} \sigma_j^h(\mathbf{x}) &= 1, & \mathbf{x} \in \Lambda_j^h \\ &= -|\Lambda_j^h|/|\Lambda_j^{2h}|, & \mathbf{x} \in \Lambda_j^{2h} \end{aligned} \quad (53)$$

Thus σ_j^h is neutral. Let $\varphi_j^h(\mathbf{x})$ be the Coulomb potential of σ_j^h . We require that there exists a function $C(h) < \infty$ such that for all $\mathbf{x} \in \Lambda_j^{2h} \cup \Lambda_j^h \cup \Lambda_j$

$$|\varphi_j^h(\mathbf{x})| < C(h) \quad (54)$$

and that

$$\lim_{h \rightarrow 0} C(h) = 0$$

(iii) Let E_j^h be the Coulomb self-energy of the double layer σ_j^h . We require that

$$\lim_{j \rightarrow \infty} |\Lambda_j|^{-1} E_j^h = 0 \quad (55)$$

Conditions (i) and (ii) obviously imply (iii) since

$$E_j^h \leq \frac{1}{2} [\sup_{\mathbf{x}} |\sigma_j^h(\mathbf{x})| |\varphi_j^h(\mathbf{x})|] [|\Lambda_j^{2h}| + |\Lambda_j^h|]$$

and

$$[|\Lambda_j^{2h}| + |\Lambda_j^h|] / |\Lambda_j| \rightarrow 0$$

by the Van Hove limit.

(iv) Define $\tilde{\Lambda}_j^h$, $\tilde{\Lambda}_j^{2h}$, $\tilde{\sigma}_j^h(x)$, and $\tilde{E}_j^h$ similarly to the above except $\mathbf{x} \notin \Lambda_j$ (resp. $\mathbf{x} \notin \Lambda_j^h \cup \Lambda_j$) is replaced by $\mathbf{x} \in \Lambda_j$ (resp. $\mathbf{x} \in \Lambda_j \setminus \Lambda_j^h$). That is, the double layer $\tilde{\sigma}_j^h$ is now inside Λ_j . We require that

$$\lim_{j \rightarrow \infty} |\Lambda_j|^{-1} \tilde{E}_j^h \rightarrow 0 \quad (56)$$

We do not require that the analogs of (i) and (ii) hold.

It is clear that for any reasonable sequence of domains, such as cubes or ellipsoids, these conditions will be satisfied. We shall not attempt to determine geometric conditions on the Λ_j so that (i)–(iv) hold.

Let Λ_j contain j particles. As in the classical case, we derive a lower bound for $Z(\Lambda_j)$. The kinetic energy for D_j can be handled as in Section 5.1. The only essential difference from inequality (15) is that we have to add the self-energy of the double layer $\tilde{\sigma}_j^h$ inside Λ_j and that of the double layers of the balls in the packing of Λ_j . Call this latter quantity $W(j)$. Thus, on the right side of the inequality we must add $-\beta \tilde{E}_j^h - \beta W(j)$. Using condition (iv), we have that

$$\liminf g_j \geq g(\beta, \rho) \quad (57)$$

An upper bound for $Z(\Lambda_j)$ is also obtained as in the classical case (18). For the domain D_j we choose a vector state ψ and have to compute

$$E(\psi) = \langle H_K \rangle + \langle U(D_j, \Lambda_j) \rangle \quad (58)$$

The kinetic energy part of $\langle H_K \rangle$ can be handled as in Section 5.1. If M_j is the number of particles in D_j , we have to find M_j points in $D_j'' = D_j -$

$\{\bigcup S^i \cup \Lambda_j^h\}$ in such a way that $E(\psi)$ is not too large. The novel feature is that $U(D_j, \Lambda_j)$ involves the additional term

$$\sum_{i=1}^{M_j} w(\mathbf{x}_i) \quad \text{where} \quad w(\mathbf{x}) = \int_{\Lambda_j} \rho_j(\mathbf{y}) |\mathbf{x} - \mathbf{y}|^{-1} d\mathbf{y}$$

and $\rho_j(\mathbf{y})$ is the average charge distribution (including the background) in Λ_j in the canonical distribution. Although Λ_j is neutral, $w \neq 0$ because Λ_j is not spherical. To find these M_j points we again average over all allowed configuration in D_j' .

The self-energy of the double layers S^i and T^i is small for the same reason as in Section 5.1. The problem then reduces to computing E_j^h as defined in condition (iii), together with the energy of the charge distribution σ_j^h in the potential w . Condition (iii) states that $|\Lambda_j|^{-1} E_j^h \rightarrow 0$, so we can ignore it. The latter contribution, Δ_j , can be bounded as follows:

$$\begin{aligned} |\Delta_j| &= \left| \int w(\mathbf{x}) \rho \sigma_j^h(\mathbf{x}) \right| \\ &= \left| \int_{\Lambda_j} d\mathbf{x} \int d\mathbf{y} \rho_j(\mathbf{x}) |\mathbf{x} - \mathbf{y}|^{-1} \rho \sigma_j^h(\mathbf{x}) \right| \\ &= \rho \left| \int_{\Lambda_j} d\mathbf{x} \rho_j(\mathbf{x}) \varphi_j^h(\mathbf{x}) \right| \leq \rho C(h) \int_{\Lambda_j} d\mathbf{x} |\rho_j(\mathbf{x})| \\ &\leq \rho C(h) \int_{\Lambda_j} |\rho_+(\mathbf{x})| + |\rho_-(\mathbf{x})| \leq 2j\rho C(h) \end{aligned} \quad (59)$$

where $\rho_+ = \rho$ is the background charge and $\rho_-(\mathbf{x})$ is the average particle charge distribution.

Now we divide by $|\Lambda_j| = j$ and let $j \rightarrow \infty$. For each fixed h we obtain

$$\limsup_{j \rightarrow \infty} g_j \leq g(\beta, \rho) + \beta \rho C(h) \quad (60)$$

Since h is arbitrary, we can now let $h \rightarrow 0$ and, recalling condition (ii), obtain

$$\lim_{j \rightarrow \infty} g_j = g(\beta, \rho) \quad (61)$$

which is the desired result.

5.3. Scaling Relations

In Section 3.3 we showed that $g(\beta, \rho) = \rho(1 - \ln \rho) + \rho \bar{g}(\beta \rho^{1/3})$. Such a simple relation will not hold quantum mechanically. To obtain a similar result quantum mechanically, we have to add another parameter; the simplest

is α , the square of the electric charge. Thus $H = K + U \rightarrow K + \alpha U$, where K is the kinetic energy operator.

We make a scale change which now involves α :

$$\rho' = \rho\eta^3, \quad \beta' = \beta\eta^{-2}, \quad \alpha' = \alpha\eta, \quad \Lambda_j' = \eta^{-1}\Lambda_j \quad (62)$$

Then, as in Section 3.3 [Eq. (22)],

$$|\Lambda_j|g(\beta, \rho, \alpha; \Lambda_j) = |\Lambda_j'|g(\beta\eta^{-2}, \rho\eta^3, \alpha\eta; \Lambda_j') \quad (63)$$

Again, choosing $\eta = \rho^{-1/3}$, and taking the limit $j \rightarrow \infty$, we obtain

$$g(\beta, \rho, \alpha) = \rho g(\beta\rho^{2/3}, 1, \alpha\rho^{-1/3}) \quad (64)$$

This equation tells us nothing that we did not know before, i.e., α is an inessential parameter. But it does tell us something important about the continuity with respect to ρ . Define $\gamma = \beta\alpha$. Then

$$\ln Z = \ln \text{Tr} \exp(-\beta K - \gamma U) \quad (65)$$

is a *jointly* convex function of (β, γ) for $\beta > 0$. Thus, when $\gamma > 0$, the thermodynamic limit $g(\beta, \rho, \gamma\beta^{-1})$ is convex, and hence continuous, in (β, γ) . Hence the function $g(x, 1, y)$ is continuous in (x, y) when $x, y > 0$. Therefore $g(\beta, \rho, \alpha)$ is continuous in ρ for $\rho > 0$.

5.4. Properties of the Thermodynamic Limit and Related Questions

The results given for the classical case in Sections 3.4–3.6 and 4 hold for the quantum case. The conclusions of Section 3.7 have to be modified. In summary one has:

- (i) Uniformity and continuity of the limit.
- (ii) Unusual behavior of the pressure and compressibility.
- (iii) Equivalence of canonical and microcanonical ensembles. The existence of the thermodynamic limit of the microcanonical ensemble includes as a special case the existence of the limiting ground state energy per unit volume. This is also true classically.
- (iv) Existence of the grand canonical pressure even for strictly neutral systems because for large ρ , the quantum kinetic energy, which behaves like $\rho^{5/3}$, will dominate the electrostatic $-\rho^{4/3}$ term. We shall not prove this statement since the lack of convexity in ρ prevents the grand canonical ensemble from being equivalent to the canonical ensemble.
- (v) The possibility of adding tempered potentials, with the same caveat as in Section 4.

APPENDIX. A LOWER BOUND FOR THE CLASSICAL AND QUANTUM MECHANICAL GROUND-STATE ENERGY

Consider a bounded, measurable set Λ with a uniform charge density ρ and N point particles of charge -1 . We do not assume that Λ is spherical, that the points are constrained to lie in Λ , or that the total charge is zero.

To find a lower bound for the total electrostatic energy we use an idea of Onsager⁽¹²⁾ to replace point charge distributions by charges smeared around the initial points. In fact one can show, by taking functional derivatives, that the best smearing is a uniform charge distribution inside a ball of radius a .

We define

- U_{BB} the self-energy of the background;
- U_i the interaction energy of the particle i at position $\mathbf{x}_i$ with the background;
- U_{ij} the interaction of two particles at positions $\mathbf{x}_i$ and $\mathbf{x}_j$;
- $\hat{U}_{ij}(a)$ the interaction (or twice the self-energy when $i = j$) of balls of total charge -1 with centers $\mathbf{x}_i$ and $\mathbf{x}_j$;
- $\hat{U}_i(a)$ the interaction of such a ball with the background.

Then, with $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$

$$U(\mathbf{X}) = U_{BB} + \sum_{i=1}^N U_i + \sum_{i < j} U_{ij}$$

$$U(\mathbf{X}) = U_{BB} + \sum_{i=1}^N \hat{U}_i + \frac{1}{2} \sum_{i,j} \hat{U}_{ij} \quad (\alpha)$$

$$+ \sum_i (U_i - \hat{U}_i) \quad (\beta)$$

$$- \frac{1}{2} \sum_i \hat{U}_{ii} \quad (\gamma)$$

$$+ \sum_{i < j} (U_{ij} - \hat{U}_{ij}) \quad (\delta)$$

Let us consider these terms individually:

The (α) terms are evidently positive, being the total electrostatic energy of the background charge and the charged balls.

In (δ) , a term $U_{ij} - \hat{U}_{ij}$ is zero if the two balls do not overlap by Newton's theorem. For overlapping balls, a simple calculation shows that $U_{ij} - \hat{U}_{ij} > 0$. Thus (δ) is positive.

For (β) we calculate

$$U_i - \hat{U}_i \geq -(2\pi/5)\rho a^2 \quad (\text{A.1})$$

The above is an equality whenever the ball lies completely in Λ .

For (γ)

$$\hat{U}_{ii} = 6/(5a) \quad (\text{A.2})$$

Our lower bound is $(\beta) + (\gamma)$, and the best bound is obtained when a is

$$a_{\max} = \rho^{-1/3}(3/4\pi)^{1/3} = r_s \quad (\text{A.3})$$

With this value we obtain

$$U(\mathbf{X}) \geq -0.9Nr_s^{-1} = -0.9N\rho^{1/3}(4\pi/3)^{1/3} \quad (\text{A.4})$$

for all $\mathbf{X}$.

For the quantum mechanical case, when the particles are spin- $\frac{1}{2}$ fermions, we consider a sequence of domains $\{\Lambda_j\}$ which tend to infinity in the sense of Van Hove and we constrain the particles to lie in Λ_j . We also suppose that $\rho|\Lambda_j| = j$, the number of particles, although this neutrality restriction is not essential in what follows.

Let

$$E_j = \inf_{\psi} \langle \psi | H | \psi \rangle |\Lambda_j|^{-1} \geq \inf_{\psi} |\Lambda_j|^{-1} \langle \psi | K | \psi \rangle + \inf_{\psi} |\Lambda_j|^{-1} \langle \psi | U | \psi \rangle$$

For $m = 1$ and j large

$$\begin{aligned} \inf \langle \psi, K\psi \rangle &\approx j\rho^{2/3}(6/5)(3\pi^2)^{2/3} = jr_s^{-2}(6/5)(9\pi/4)^{2/3} \\ &= 2.21jr_s^{-2} \text{ Ry} \end{aligned} \quad (\text{A.5})$$

Therefore, to leading order in j ,

$$j^{-1}H \geq r_s^{-2}(2.21 - 0.45r_s) \text{ Ry} \quad (\text{A.6})$$

It makes sense here, unlike the situation for the classical or the Bose problem, to ask for the r_s that gives the lowest ground-state energy per particle. We obtain

$$r_s = 9.82 \quad (\text{A.7})$$

and

$$j^{-1}H \geq -0.0229 \text{ Ry} \quad (\text{A.8})$$

ACKNOWLEDGMENT

We would like to thank Prof. W. Thirring for stimulating discussions and for his hospitality at the Institute for Theoretical Physics, University of Vienna, where this work was started.

REFERENCES

1. E. P. Wigner, *Trans. Faraday Soc. (London)* **34**: 678 (1938).
2. R. J. Baxter, *Proc. Camb. Phil. Soc.* **59**: 779 (1963).
3. H. Kunz, *Ann. Phys. (N.Y.)* **85**: 303 (1974).

4. H. J. Brascamp and E. H. Lieb, Some inequalities for Gaussian measures and the long-range order of the one-dimensional plasma, in *Proceedings of the Conference on Functional Integration*, London (1974), to be published.
5. E. H. Lieb and J. L. Lebowitz, *Adv. in Math.* **9**:316 (1972).
6. J. L. Lebowitz and E. H. Lieb, *Phys. Rev. Lett.* **22**:631 (1969).
7. O. Penrose and E. R. Smith, *Commun. Math. Phys.* **26**:53 (1972).
8. F. J. Dyson and A. Lenard, *J. Math. Phys.* **8**:423 (1967).
9. H. Narnhofer and W. Thirring, Convexity properties of Coulomb systems, *Acta Phys. Austr.* (1975), to be published.
10. A. Lenard and F. J. Dyson, *J. Math. Phys.* **9**:698 (1968).
11. P. Federbush, *J. Math. Phys.* **16**:706 (1975).
12. L. Onsager, *J. Phys. Chem.* **43**:189 (1939).
13. R. A. Goldwell-Horstall and A. A. Maradudin, *J. Math. Phys.* **1**:395 (1960).

Erratum

The Thermodynamic Limit for Jellium¹

Elliott H. Lieb² and Heide Narnhofer³

Received September 22, 1975

On p. 309, Eq. (A.5) should read

$$\begin{aligned}\inf\langle\psi, K\psi\rangle &\approx j\rho^{2/3}(3/5)(3\pi^2)^{2/3} = jr_s^{-2}(3/5)(9\pi/4)^{2/3} \\ &= 2.21jr_s^{-2} \text{ Ry}\end{aligned}\tag{A.5}$$

Accordingly, the resulting expressions should read as follows:

$$j^{-1}H \geq r_s^{-2}(2.21 - 1.80r_s) \text{ Ry}\tag{A.6}$$

$$r_s = 2.46\tag{A.7}$$

$$j^{-1}H \geq -0.367 \text{ Ry}\tag{A.8}$$

¹ This paper appeared in *J. Stat. Phys.* **12**(4):291 (1975).

² Department of Mathematics and Physics, Princeton University, Princeton, New Jersey.

³ Institute for Theoretical Physics, University of Vienna, Vienna, Austria.

Part VII

Quantum Electrodynamics

Self-Energy of Electrons in Non-Perturbative QED

Elliott H. Lieb*

Departments of Mathematics and Physics

Princeton University, Princeton, New Jersey 08544-0708

Michael Loss†

School of Mathematics

Georgia Institute of Technology, Atlanta, Georgia 30332-0160 ‡§

Abstract

Various models of charged particles interacting with a quantized, ultraviolet cutoff radiation field (but not with each other) are investigated. Upper and lower bounds are found for the self- or ground state-energies without mass renormalization. For N fermions the bounds are proportional to N , but for bosons they are sublinear, which implies ‘binding’, and hence that ‘free’ bosons are never free. Both ‘relativistic’ and non-relativistic kinematics are considered. Our bounds are non-perturbative and differ significantly from the predictions of perturbation theory.

1 Introduction

Quantum electrodynamics (QED), the theory of electrons interacting with photons (at least for small energies) is one of the great successes of physics. Among its major achievements is the explanation of the Lamb shift and the anomalous magnetic moment of the electron. Nevertheless, its computations, which are entirely based on perturbation theory, created some uneasiness among the practitioners. The occurrence of

*Supported in part by NSF Grant PHY-98 20650.

†Supported in part by NSF Grant DMS-95 00840.

‡©1999 by the authors. Reproduction of this work, in its entirety, by any means, is permitted for non-commercial purposes.

§This work was presented by E.H.L. at the University of Alabama, Birmingham - Georgia Tech International Conference on Differential Equations and Mathematical Physics, Birmingham, March 15-19, 1999 and first appears in the proceedings of that conference published by the American Mathematical Society and International Press, “Studies in Advanced Mathematics”, vol 16, pp. 279-293 (2000). A second version, with some typos corrected, appears in the proceedings of ‘Conference Moshe Flato’ in Dijon in September, 1999, published by Kluwer, “Mathematical Physics Studies”, vol.21, pp. 327-344 (2000). The present version, containing additional corrections, was prepared for the third edition of the *Selecta*.

infinities was and is especially vexing. Moreover, a truly nontrivial, 3+1-dimensional example of a relativistically invariant field theory has not yet been achieved.

There are, however, unresolved issues at a much earlier stage of QED that hark back to black-body radiation, the simplest and historically first problem involving the interaction of matter with radiation. The conceptual problems stemming from black-body radiation were partly resolved by quantum mechanics, i.e., by the non-relativistic Schrödinger equation, which is, undoubtedly, one of the most successful of theories, for it describes matter at low energies almost completely. It is mathematically consistent and there are techniques available to compute relevant quantities. Moreover, it allows us to explain certain facts about bulk matter such as its stability, its extensivity, and the existence of thermodynamic functions. What has not been as successful, so far, is the incorporation of radiation phenomena, the very problem quantum mechanics set out to explain.

It ought to be possible to find a mathematically consistent theory, free of infinities, that describes the interaction of non-relativistic matter with radiation at moderate energies, such as atomic binding energies. It should not be necessary, as some physicists believe, to view QED as a low energy part of a consistent high energy theory.

From such a theory one could learn a number of things that have not been explained rigorously. i) The decay of excited states in atoms. This problem has been investigated in some ultraviolet cutoff models in [BFS] and in a massive photon model in [OY]. See also the review of Hogreve [H]. ii) Non-relativistic QED could be a playground for truly non-perturbative calculations and it could shed light on renormalization procedures. In fact, this was the route historically taken by Kramers that led to the renormalization program of Dyson, Feynman, Schwinger and Tomonaga. iii) Last but not least, one could formulate and answer the problems of stability of bulk matter interacting with the radiation field.

It has been proved in [F],[LLS] that stability of non-relativistic matter (with the Pauli Hamiltonian) interacting with classical magnetic fields holds provided that the fine-structure constant, $\alpha = e^2/\hbar c$, is small enough. It is certain, that the intricacies and difficulties of this classical field model will persist and presumably magnify in QED.

The same may be expected from a relativistic QED since replacing the Pauli Hamiltonian by a Dirac operator leads to a similar requirement on α [LSS]. Indeed, stability of matter in this model (the Brown-Ravenhall model) requires that the electron (positron) be defined in terms of the positive (negative) spectral subspace of the Dirac operator *with* the magnetic vector potential $A(x)$, instead of the free Dirac operator *without* $A(x)$. This observation, that perturbation theory, if there is one, must start from the dressed electrons rather than the electrons unclothed by its magnetic field, might ultimately be important in a non-perturbative QED.

The first, humble, step is to understand electrons that interact with the radiation field but which are free otherwise. In order for this model to make sense an ultraviolet cutoff has to be imposed that limits the energy of photon modes. The simplest question, which is the one we address in this paper, is the behavior of the self-energy of the electron as the cutoff tends to infinity (with the bare mass of the electron fixed). The self-energy of the electron diverges as the cutoff tends to infinity and it has to be subtracted for each electron in any interacting theory. The total energy will still depend strongly on the cutoff because of the interactions. This dependence

will, hopefully, enter through an effective mass which will be set equal to the physical mass (mass renormalization). The resulting theory should be essentially Schrödinger's mechanics, but slightly modified by so-called radiative corrections.

Lest the reader think that the self-energy problem is just a mathematical exercise, consideration of the many-body problem will provide a counterexample. Imagine N charged bosons interacting with the radiation field, but neglect any interaction among them such as the Coulomb repulsion. We say that these particles bind if the energy of the combined particles is less than the energy of infinitely separated particles. As we shall show, charged bosons indeed bind and they do it in such a massive way that it will be very likely that this cannot be overcome by the Coulomb repulsion. In particular, the energy of a charged many-boson system is *not* extensive, and from this perspective it is fortunate that stable, charged bosons do not exist in nature.

The situation is very different for fermions. We are not able to show that they do not bind, but we can show — and this is one of the main results of our paper — that the self-energy is extensive, i.e., bounded above and below by a constant times N .

We thus have strong evidence that there is *no* consistent description of a system of stable charged relativistic or non-relativistic bosons interacting with the radiation field, while the Pauli exclusion principle, on the other hand, is able to prevent the above mentioned pathology.

In the remainder of the section we explain our notation and state the results. In the subsequent sections we sketch the proof of some of them but for details we refer the reader to [LL].

We measure the energy in units of mc^2 where m is the bare mass of the electron, the length in units of the Compton wave length $\ell_C = \hbar/mc$ of the bare electron. We further choose $\ell_C^{-1}\sqrt{\hbar c}$ as the unit for the vector potential A and $\ell_C^{-2}\sqrt{\hbar c}$ as the unit for the magnetic field B . As usual, $\alpha = e^2/\hbar c \approx 1/137.04$ is the fine structure constant.

In the expression below, $A(x)$ denotes an ultraviolet cutoff radiation field localized in a box $L \times L \times L$ with volume $V = L^3$,

$$A(x) = \frac{1}{\sqrt{2V}} \sum_{|k| < \Lambda} \sum_{\lambda=1,2} \frac{1}{\sqrt{|k|}} \varepsilon_\lambda(k) [a_\lambda(k) e^{ix \cdot k} + a_\lambda^*(k) e^{-ix \cdot k}] . \quad (1.1)$$

The index $k = 2\pi n/L$ where $n \in \mathbb{Z}^3$, and the word cutoff refers to the restriction to all values of k with $|k| < \Lambda$.

The vectors $\varepsilon_\lambda(k)$ are the polarization vectors and are normalized in such a way that

$$\varepsilon_\lambda(k) \cdot \varepsilon_{\lambda'}(k) = \delta_{\lambda,\lambda'} , \quad \varepsilon_\lambda(k) \cdot k = 0 . \quad (1.2)$$

The operators $a_\lambda(k)$ and $a_\lambda^*(k)$ satisfy the commutation relation

$$[a_\lambda(k), a_{\lambda'}^*(k')] = \delta_{\lambda,\lambda'} \delta(k, k') , \quad (1.3)$$

while all others commute with each other.

The energy of the radiation field can now be conveniently written as

$$H_f = \sum_{|k| < \Lambda} \sum_{\lambda=1,2} |k| a_\lambda^*(k) a_\lambda(k) . \quad (1.4)$$

These operators act on the Hilbert space generated by the polynomials in $a_\lambda^*(k)$ acting on the vacuum $|0\rangle$.

The self energy of (one or more) particles is the **ground state energy** of the Hamiltonian

$$H = \text{kinetic energy} + H_f , \quad (1.5)$$

where, as usual, the ground state energy of H is defined to be

$$E_0 = \inf_{\Psi} \frac{\langle \Psi, H \Psi \rangle}{\langle \Psi, \Psi \rangle} . \quad (1.6)$$

Typically, in the inquiry about the self-energy problem, i.e., the problem of computing the self-energy for fixed, albeit small, α and for large Λ , one proceeds via perturbation theory. First order perturbation theory will predict an energy of the order of $\alpha\Lambda^2$, and a higher order power counting argument confirms the asymptotically large Λ dependence of that calculation. Our theorems below show that the predictions of perturbation theory for the self-energy problem are wrong, if one is interested in the large Λ asymptotics of the energy. If perturbation theory works at all, then it works only for a range of α that vanishes as Λ increases. In fact we deduce from the upper bound in Theorem 1.1 that the size of this range shrinks at least as $\Lambda^{-2/5}$.

All the theorems below are asymptotic statements for *large* Λ and for *fixed* α , and all the constants are *independent of the volume*. For actual bounds we refer the reader to [LL]. The first result concerns the self energy of a nonrelativistic electron interacting with the radiation field. The Hamiltonian is given by

$$H = \frac{1}{2}(p + \sqrt{\alpha}A(x))^2 + H_f , \quad (1.7)$$

where $p = -i\nabla$ and acts on $L^2(\mathbb{R}^3) \otimes \mathcal{F}$, where $\mathcal{F}$ denotes the photon Fock space.

Theorem 1.1 *The ground state energy, E_0 , of the operator (1.7) satisfies the asymptotic bounds (with positive constants C_1, C_2)*

$$C_1\alpha^{1/2}\Lambda^{3/2} < E_0 < C_2\alpha^{2/7}\Lambda^{12/7} \quad (1.8)$$

We do not know how to get upper and lower bounds that are of the same order in Λ , but we suspect that $\Lambda^{12/7}$ is the right exponent. This is supported by the following theorem in which the $p \cdot A$ term is omitted.

Theorem 1.2 *The ground state energy E_0 of the operator*

$$\frac{1}{2} [p^2 + \alpha A(x)^2] + H_f \quad (1.9)$$

satisfies the asymptotic bounds (with positive constants C_1, C_2)

$$C_1\alpha^{2/7}\Lambda^{12/7} \leq E_0 \leq C_2\alpha^{2/7}\Lambda^{12/7} . \quad (1.10)$$

While these results are not of direct physical relevance (since E_0 is not observable), the many-body problem is of importance since it reveals a dramatic difference between bosons and fermions.

Theorem 1.3 *The ground state energy of N bosons, $E_0^{boson}(N)$, with Hamiltonian*

$$H(N) = \sum_{j=1}^N \frac{1}{2} (p_j + \sqrt{\alpha} A(x_j))^2 + H_f \quad (1.11)$$

satisfies the asymptotic bounds (with positive constants C_1, C_2)

$$C_1 \sqrt{N} \sqrt{\alpha} \Lambda^{3/2} \leq E_0^{boson}(N) \leq C_2 N^{5/7} \alpha^{2/7} \Lambda^{12/7} . \quad (1.12)$$

Thus, the energy $E_0^{boson}(N)$ is not extensive, i.e., it costs a huge energy to separate charged bosons (in the absence of their Coulomb repulsion). This has to be contrasted with the next theorem about fermions. The Hamiltonian is the same as before but it acts on the Hilbert space

$$\mathcal{F} \otimes \wedge_{j=1}^N L^2(\mathbb{R}^3; \mathbb{C}^2) , \quad (1.13)$$

where the wedge product indicates that the antisymmetric tensor product is taken.

Theorem 1.4 *The ground state energy, $E_0^{fermion}(N)$, of N charged fermions interacting with the radiation field satisfies the asymptotic bounds (with positive constants C_1, C_2)*

$$C_1 \alpha^{1/2} \Lambda^{3/2} N \leq E_0^{fermion}(N) \leq C_2 \alpha^{2/7} \Lambda^{12/7} N . \quad (1.14)$$

The “relativistic” kinetic energy for an electron is

$$T^{\text{rel}} = |p + \sqrt{\alpha} A(x)| = \sqrt{[p + \sqrt{\alpha} A(x)]^2} \quad (1.15)$$

with $p = -i\nabla$. (Really, we should take $\sqrt{[p + \sqrt{\alpha} A(x)]^2 + 1}$, but since $x < \sqrt{x^2 + 1} < x + 1$, the difference is bounded by N .)

Consider, first, the $N = 1$ body problem with the Hamiltonian

$$H = T^{\text{rel}} + H_f . \quad (1.16)$$

By simple length scaling (with a simultaneous scaling of the volume V) we easily see that $E_0 = \inf \text{spec}(H) = C\Lambda$. Our goal here is to show that the constant, C , is strictly positive and to give an effective lower bound for it. But we would like to do more, namely investigate the dependence of this constant on α . We also want to show, later on, that for N fermions the energy is bounded below by a positive constant times $N\Lambda$. Our proof will contain some novel — even bizarre — features.

Theorem 1.5 *For the Hamiltonian in 1.16 there are positive constants, C, C', C'' and α_0 , independent of α and N , and such that*

$$\begin{aligned} E_0 &\leq C\sqrt{\alpha}\Lambda \\ E_0 &\geq C'\sqrt{\alpha}\Lambda \text{ for } \alpha \leq \alpha_0 \\ E_0 &\geq C''\Lambda \text{ for } \alpha \geq \alpha_0 . \end{aligned}$$

The generalization of this to N fermions is similar to the nonrelativistic generalization, except that the power of Λ is the same on both sides of the inequalities.

Theorem 1.6 *For N fermions with Hamiltonian*

$$H_N = \sum_{i=1}^N T^{\text{rel}}(x_i) + H_f$$

there are positive constants C, C', C'' , and α_0 , independent of α and N , such that

$$\begin{aligned} E_0 &\leq CN\sqrt{\alpha}\Lambda \\ E_0 &\geq C'N\sqrt{\alpha}\Lambda \text{ for } \alpha \leq \alpha_0 \\ E_0 &\geq C''N\Lambda \text{ for } \alpha \geq \alpha_0 . \end{aligned} \tag{1.17}$$

We close this introduction by mentioning one last result about the Pauli-operator. The kinetic energy expression is given by

$$T^{\text{Pauli}} = [\sigma \cdot (p + \sqrt{\alpha}A(x))]^2 = (p + \sqrt{\alpha}A(x))^2 + \sqrt{\alpha} \sigma \cdot B(x) . \tag{1.18}$$

where σ denotes the vector consisting of the Pauli matrices. Observe that this term automatically accounts for the spin-field interaction. Our result for the self energy of a Pauli electron is the following.

Theorem 1.7 *For the Hamiltonian with Pauli kinetic energy,*

$$\frac{1}{2}[\sigma \cdot (p + \sqrt{\alpha}A(x))]^2 + H_f, \tag{1.19}$$

there are positive constants C, C', C'' , and α_0 , such that the ground state energy satisfies the asymptotic bounds

$$\begin{aligned} E_0 &\leq C_3\sqrt{\alpha}\Lambda^{3/2} \\ E_0 &\geq C_1\alpha\Lambda \text{ for } \alpha \leq \alpha_0 \\ E_0 &\geq C_2\alpha^{1/3}\Lambda \text{ for } \alpha \geq \alpha_0 . \end{aligned} \tag{1.20}$$

For N fermions, the bounds above are multiplied by N (and the constants are changed).

For the details of the proof, we refer the reader to [LL]. We believe that the upper bound is closer to the truth since the main contributions to the self energy should come from the fluctuations of the A^2 term.

Theorem 1.7 has the following consequence for stability of matter interacting with quantized fields. It was shown in [LLS] that a system of electrons and nuclei interacting with Coulomb forces, with the Pauli kinetic energy for the electrons and with a *classical* magnetic field energy is stable (i.e., the ground state energy is bounded below by N) if and only if α is small enough. In [BFG],[FFG] this result was extended to *quantized*, ultraviolet cutoff magnetic fields (as here). Among other things, it was shown in [FFG] that the ground state energy, E_0 , of the electrons and nuclei problem is bounded below by $-\alpha^2 \Lambda N$ for small α . Theorem 1.7 implies, as a corollary, that for small α the *total energy* (including Coulomb energies) is bounded below by $+\alpha \Lambda N$. In other words, among the three components of energy (kinetic, field and Coulomb), the first two overwhelm the third — for small α , at least.

All of these statements are true without mass renormalization and the situation could conceivably be more dramatic when the mass is renormalized. In any case, the true physical questions concern energy differences, and this question remains to be addressed.

2 Non-Relativistic Energy Bounds

Theorem 1.1: We sketch a proof of Theorem 1.1. It is clear by taking the state $V^{-1/2} \otimes |0\rangle$ (where $|0\rangle$ denotes the photon vacuum) that the ground state energy is bounded above by $(\text{const})\alpha\Lambda^2$, which is the same result one gets from perturbation theory. Since the field energy in this state vanishes, such a computation ignores the tradeoff between the kinetic energy of the electron and the field energy. Thus, it is important to quantify this tradeoff. The main idea is to estimate the field energy in terms of selected modes.

Consider the operators (field modes), parametrized by $y \in \mathbb{R}^3$,

$$L(y) = \frac{1}{\sqrt{2V}} \sum_{|k| < \Lambda, \lambda} \sqrt{|k|} a_\lambda(k) \bar{v}_\lambda(k) e^{ik \cdot y}, \quad (2.1)$$

with some arbitrary complex coefficients $v_\lambda(k)$. The following lemma is elementary

Lemma 2.1

$$H_f \geq \int w(y) L^*(y) L(y) dy \quad (2.2)$$

provided that the functions $v_\lambda(k)$ and w are chosen such that, as matrices,

$$|k| \delta_{\lambda', \lambda} \delta(k', k) \geq \frac{1}{2V} \sqrt{|k|} \bar{v}_\lambda(k) \hat{w}(k - k') v_{\lambda'}(k') \sqrt{|k'|}, \quad (2.3)$$

or equivalently, that

$$\sum_{|k| < \Lambda, \lambda} \frac{|f_\lambda(k)|^2}{|v_\lambda(k)|^2} \geq \sum_{|k|, |k'| < \Lambda, \lambda, \lambda'} \frac{1}{2V} \overline{f_\lambda(k)} f_{\lambda'}(k') \hat{w}(k - k') \quad (2.4)$$

for all $f_\lambda(k)$, where $\hat{w}(k) = \int e^{ik \cdot x} w(x) dx$.

For the proof, one simply notes that both sides of (2.2) are quadratic forms in the creation and annihilation operators, and hence (2.3) and (2.4) are necessary and sufficient conditions for (2.2) to be true. ■

Corollary 2.2 *Assuming (2.2), the following two inequalities hold.*

$$H_f \geq -\frac{1}{4V} \sum_{|k| < \Lambda, \lambda} |k| |v_\lambda(k)|^2 \int |w(y)| dy + \frac{1}{4} \left\{ \int w(y) (L(y) + L^*(y))^2 dy - \int w(y) (L(y) - L^*(y))^2 dy \right\}. \quad (2.5)$$

To prove this, note that

$$L^*L = LL^* - \frac{1}{2V} \sum_{|k| < \Lambda, \lambda} |k| |v_\lambda(k)|^2, \quad (2.6)$$

and, quite generally for operators,

$$\pm(LL + L^*L^*) \leq L^*L + LL^*. \quad (2.7)$$

■

Returning to the proof of Theorem 1.1 we start with the lower bound. Denote by

$$\Pi(x) = \frac{-i}{\sqrt{2V}} \sum_{|k| < \Lambda, \lambda} \sqrt{|k|} \varepsilon_\lambda(k) (a_\lambda(k) e^{ik \cdot x} - a_\lambda^*(k) e^{-ik \cdot x}). \quad (2.8)$$

This operator is canonically conjugate to $A(x)$ in the sense that we have the commutation relations

$$i[\Pi_i(x), A_j(x)] = \delta_{i,j} \frac{1}{(3\pi)^2} \Lambda^3. \quad (2.9)$$

For our calculation, it is important to note that

$$\operatorname{div} \Pi(x) = 0. \quad (2.10)$$

Hence from (2.9) and (2.10) we get that

$$\sum_{j=1}^3 [p_j + \sqrt{\alpha} A_j(x), \Pi_j(x)] = \sqrt{\alpha} \frac{i}{3\pi^2} \Lambda^3. \quad (2.11)$$

The inequality

$$\frac{1}{2} (p + \sqrt{\alpha} A(x))^2 + 2a^2 \Pi(x)^2 \geq -ai \sum_{j=1}^3 [p_j + \sqrt{\alpha} A_j(x), \Pi_j(x)], \quad (2.12)$$

valid for all positive numbers a , yields

$$\frac{1}{2}(p + \sqrt{\alpha}A(x))^2 + H_f \geq a\sqrt{\alpha}\frac{1}{3\pi^2}\Lambda^3 + H_f - 2a^2\Pi(x)^2. \quad (2.13)$$

We choose

$$v_\lambda(k) = 3\pi\Lambda^{-3/2}\varepsilon_{\lambda,j}(k) \quad (2.14)$$

(where $j = 1, 2, 3$ indexes the polarization vector component) and

$$w(y) = 12\beta \delta(x - y). \quad (2.15)$$

Eqn. (2.3) is satisfied if and only if $\beta \leq 3\pi^2\Lambda^{-3}/2$, in which case Corollary 2.2 yields an inequality which, after summing on j , is

$$H_f \geq \beta \Pi(x)^2 - \frac{\beta}{(2\pi)^2}\Lambda^4. \quad (2.16)$$

In eqn. (2.13) we choose $2a^2 = \beta = 3\pi^2\Lambda^{-3}/2$, i.e., $a = (\sqrt{3}\pi)/(2\Lambda^{3/2})$. This yields the *lower bound*

$$H \geq \frac{1}{2\pi}\sqrt{\frac{\alpha}{3}}\Lambda^{3/2} - \frac{3}{8}\Lambda. \quad (2.17)$$

The idea of using a commutator, as in (2.12), (2.13) to estimate the ground state energy, goes back to the study of the polaron [LY].

For the upper bound we take a simple trial function of the form

$$\phi(x) \otimes \Psi \quad (2.18)$$

where $\Psi \in \mathcal{F}$ is normalized and $\phi(x)$ is a *real* function normalized in $L^2(\mathbb{R}^3)$. An upper bound to the energy is thus given by

$$\frac{1}{2} \int |\nabla \phi(x)|^2 dx + \frac{\alpha}{2} \int \phi(x)^2 (\Psi, A(x)^2 \Psi) dx + (\Psi, H_f \Psi). \quad (2.19)$$

It is not very difficult to see that the last two terms can be concatenated into the following expression.

$$\frac{1}{2} \int (\Psi, [\Pi(x)^2 + \alpha A(x)(-\Delta + \phi(x)^2)A(x)] \Psi) dx - \frac{1}{2} \text{Tr} \sqrt{P - \Delta P}. \quad (2.20)$$

Here, P is the projection onto the divergence free vector fields with ultraviolet cutoff Λ . This can be deduced by writing the field energy in terms of $\Pi(x)$ and $A(x)$. The first term in (2.20) is a sum of harmonic oscillators whose zero point energy is given by

$$\frac{1}{2} \text{Tr} \sqrt{P(-\Delta + \alpha \phi(x)^2)P} \quad (2.21)$$

and hence

$$\frac{1}{2} \text{Tr} \sqrt{P(-\Delta + \alpha \phi(x)^2)P} - \frac{1}{2} \text{Tr} \sqrt{P(-\Delta)P}, \quad (2.22)$$

is an exact expression for the ground state energy of (2.20). Using the operator monotonicity of the square root we get as an upper bound on (2.20)

$$\frac{1}{2} \int |\nabla \phi(x)|^2 dx + \frac{1}{2} \sqrt{\alpha} \text{Tr} \sqrt{P \phi(x)^2 P}. \quad (2.23)$$

As a trial function we use

$$\phi(x) = \text{const.} K^{-3/2} \int \left(1 - \frac{|k|}{K}\right)_+^3 e^{ik \cdot x} dk. \quad (2.24)$$

Optimizing the resulting expression over K yields the stated result. For details we refer the reader to [LL]. ■

It is natural to ask, how good this upper bound is. If we neglect the cross terms in $(p + A)^2$, i.e., we replace the kinetic energy by $p^2 + \alpha A(x)^2$, then we have Theorem 1.2, which we prove next.

Theorem 1.2: The upper bound was already given in Theorem 1.1 because $\langle p \cdot A \rangle = 0$ in the state (2.18). Loosely speaking equation (2.9) expresses the Heisenberg uncertainty principle for the field operators. An uncertainty principle that is quite a bit more useful is the following.

Lemma 2.3 *The following inequality holds in the sense of quadratic forms*

$$\Pi(x)^2 \geq \frac{1}{4} \frac{1}{(3\pi)^4} \Lambda^6 \frac{1}{A(x)^2}. \quad (2.25)$$

For the proof note that $[A_j(x), A_k(y)] = 0$ and compute

$$i[\Pi(x)_j, \frac{A_j(x)}{A(x)^2}] = \frac{1}{(3\pi)^2} \Lambda^3 \left[\frac{1}{A(x)^2} - 2 \left(\frac{A_j(x)}{A(x)^2} \right)^2 \right], \quad (2.26)$$

and summing over j we obtain that

$$i \sum_{j=1}^3 [\Pi(x)_j, \frac{A_j(x)}{A(x)^2}] = \frac{1}{(3\pi)^2} \Lambda^3 \frac{1}{A(x)^2} \quad (2.27)$$

Our statement follows from the Schwarz inequality. ■

To prove Theorem 1.2 we return to Lemma 2.1 and choose $v_\lambda(k) = \varepsilon_\lambda(k)$ and $w(x)$ any function ≤ 1 . Corollary 2.2 applied to each of the 3 components of $\Pi(x)$ then yields

$$H_f \geq \frac{1}{4} \int w(x-y) \Pi(y)^2 dy - \Lambda^4 \frac{3}{8\pi^2} \int w(y) dy, \quad (2.28)$$

for every $x \in \mathbb{R}^3$. By Lemma 2.1 the right side is bounded below by

$$\Lambda^6 \int w(x-y) \frac{1}{A(y)^2} dy - \Lambda^4 \int w(y) dy, \quad (2.29)$$

and hence

$$\begin{aligned} \langle \Psi, H \Psi \rangle &\geq \frac{1}{2} \int \langle \nabla \Psi(x), \nabla \Psi(x) \rangle dx + \frac{\alpha}{2} \int \langle \Psi(x), A(x)^2 \Psi(x) \rangle dx \\ &+ \Lambda^6 \int w(x-y) \langle \Psi(y), \frac{1}{A(x)^2} \Psi(y) \rangle dy dx \\ &- \Lambda^4 \int w(y) dy \int \langle \Psi(x), \Psi(x) \rangle dx. \end{aligned} \quad (2.30)$$

By Schwarz's inequality the second and third term together are bounded below by

$$\sqrt{\frac{\alpha}{2}} \Lambda^3 \int \langle \Psi(x), \Psi(y) \rangle \frac{w(x-y)}{\sqrt{\int w(z) dz}} dx dy. \quad (2.31)$$

If we restate our bound in terms of Fourier space variables we get

$$\int \left[\frac{|p|^2}{2} + \sqrt{\frac{\alpha}{2}} \Lambda^3 \frac{\widehat{w}(p)}{\sqrt{\widehat{w}(0)}} \right] \langle \widehat{\Psi}(p), \widehat{\Psi}(p) \rangle dp - \Lambda^4 \widehat{w}(0) \int \langle \widehat{\Psi}(p), \widehat{\Psi}(p) \rangle dp. \quad (2.32)$$

Choosing the function $\widehat{w}(p)$ to be $(2\pi)^3 \Lambda^{-18/7}$ times the characteristic function of the ball of radius $\Lambda^{6/7}$, we have that $w(x) \leq 1$ and it remains to optimize (2.32) over all normalized states $\widehat{\Psi}(p)$. This is easily achieved by noting that the function

$$\frac{1}{2} |p|^2 + \sqrt{\frac{\alpha}{2}} \Lambda^3 \frac{\widehat{w}(p)}{\sqrt{\widehat{w}(0)}} \quad (2.33)$$

is everywhere larger than $\Lambda^{12/7}$. Strictly speaking, the function $w(x)$ should be positive in order for the argument that led to (2.31) to be valid. This can be achieved with a different choice of $w(x)$, like the one in (2.24), that is more complicated but does not change the argument in an essential way.

3 Non-Relativistic Many-Body Energies

A problem that has to be addressed is the energy of N particles (bosons or fermions) interacting with the radiation field. If $E_0 = E_0(1)$ is the energy of one particle (which we estimated in the preceding section) then, ideally, the energy, $E_0(N)$, of N particles (which trivially satisfies $E_N \leq N E$, since the N particles can be placed infinitely far apart) ought to be, *exactly*,

$$E_0(N) = N E_0 \quad (3.1)$$

in a correct QED. In other words, in the absence of nuclei and Coulomb potentials, there should be *no binding* caused by the field energy H_f . This is what we seem to

observe experimentally, but this important topic does not seem to have been discussed in the QED literature.

Normally, one should expect binding, for the following mathematical reason: The first particle generates a field, $A(x)$, and energy E_0 . The second particle can either try to generate a field $A(x+y)$, located very far away at y or the second particle can try to take advantage of the field $A(x)$, already generated by the first particle, and achieve an insertion energy lower than E_0 .

Indeed, this second phenomenon happens for bosons — as expected. For fermions, however, the Paul principle plays a crucial role (even in the absence of Coulomb attractions). We show that $E_0(N) \geq CNE_0$ for fermions, but we are unable to show that the universal constant $C = 1$. Even if $C < 1$, the situation could still be saved by mass renormalization, which drives the bare mass to zero as Λ increases, thereby pushing particles apart.

3.1 Bosons

Theorem 1.3; This theorem concerns the ground state energy of N charged bosons. the Hamiltonian is given by 1.11 acting on the Hilbert space of symmetric functions tensored with the photon Fock space $\mathcal{F}$. It states, basically, that $C_1\sqrt{N}\sqrt{\alpha}\Lambda^{3/2} \leq E_0^{boson}(N) \leq C_2N^{5/7}\alpha^{2/7}\Lambda^{12/7}$.

The proof follows essentially that of the one particle case. The interesting fact is that it implies *binding of charged bosons* (in the absence of the Coulomb repulsion). The binding energy is defined by

$$\Delta E(N) = E_0(N) - NE_0(1)$$

and satisfies the bounds

$$\begin{aligned} \Delta E(N) &\geq C_1\sqrt{N}\sqrt{\alpha}\Lambda^{3/2} - C_2N\alpha^{2/7}\Lambda^{12/7} \\ \Delta E(N) &\leq C_2N^{2/7}\alpha^{2/7}\Lambda^{12/7} - C_1N\sqrt{\alpha}\Lambda^{3/2} \end{aligned} \tag{3.2}$$

which can be made negative for appropriately chosen N and Λ . There will be binding for all large enough N , irrespective of the cutoff Λ . It also has to be remarked that the Coulomb repulsion will, in all likelihood, not alter this result since it has an effect on energy scales of the order of Λ and not $\Lambda^{12/7}$ or $\Lambda^{3/2}$.

3.2 Fermions

The real issue for physics is what happens with fermions. We cannot show that there is no binding but we can show that the energy is extensive as in Theorem 1.4. The Hamiltonian is the same as (1.11) but it acts on antisymmetric functions tensored with $\mathcal{F}$. (Spin can be ignored for present purposes.)

Rough sketch of the proof of Theorem 1.4.

The difficulty in proving this theorem stems from the fact that the field energy is not extensive in any obvious way.

Define $\underline{X} = (x_1, \dots, x_N)$ and define the function

$$n_j(\underline{X}, R) = \#\{x_i \neq x_j : |x_i - x_j| < R\} .$$

This function counts the number of electrons that are within a distance R of the j^{th} electron. Note that this function is not smooth, so that all the following computations have to be modified. (See [LL].) We save half of the kinetic energy and write

$$H = \frac{1}{4} \sum_{j=1}^N (p_j + \sqrt{\alpha} A(x_j))^2 + H'.$$

We apply the commutator estimate (2.11) and (2.13) to the pair

$$i[p_j + \sqrt{\alpha} A(x_j), \frac{1}{\sqrt{N_j(\underline{X}, R) + 1}} \Pi(x_j)]$$

and obtain the bound (with the caveat mentioned above), for all $\alpha > 0$,

$$H' \geq a\sqrt{\alpha}\Lambda^3 \sum_{j=1}^N \frac{1}{\sqrt{N_j(\underline{X}, R) + 1}} - a^2 \sum_{j=1}^N \frac{1}{N_j(\underline{X}, R) + 1} \Pi(x_j)^2 + H_f \quad (3.3)$$

The next two steps are somewhat nontrivial and we refer the reader to [LL]. First one notes that the modes $F(x_i)$ and $F(x_j)$ are essentially orthogonal (i.e., they commute) if $|x_i - x_j| > \Lambda^{-1}$. Ignoring the technical details of how this is implemented, the key observation is that the last two terms in (3.3) can be estimated from below by $-N\Lambda$ provided $a = \Lambda^{-3/2}$.

The next ingredient is a new Lieb-Thirring type estimate involving the function $N_j(\underline{X}, R)$. It is here and only here that the Pauli exclusion principle is invoked.

Theorem 3.1 *On the space $\wedge_{j=1}^N L^2 \mathbb{R}^3; \mathbb{C}^q$ of antisymmetric functions*

$$\sum_{j=1}^N (p_j + \sqrt{\alpha} A(x_j))^2 \geq \frac{C}{q^{2/3}} \frac{1}{R^2} \sum_{j=1}^N N_j(\underline{X}, R)^{2/3} \quad (3.4)$$

with $C \geq 0.00127$. An analogous inequality holds for the relativistic case as well:

$$\sum_{j=1}^N |p_j + \sqrt{\alpha} A(x_j)| \geq \frac{C}{q^{1/3}} \frac{1}{R} \sum_{j=1}^N N_j(\underline{X}, R)^{1/3} \quad (3.5)$$

By using the kinetic energy previously saved together with (3.3) and the previous discussion, we get

$$H \geq \sum_{j=1}^N \left\{ N_j(\underline{X}, R)^{2/3} + \sqrt{\alpha} \Lambda^{3/2} \frac{1}{\sqrt{N_j(\underline{X}, R) + 1}} \right\} - N\Lambda.$$

By minimizing over N_j the desired lower bound in Theorem (1.4) is obtained. The upper bound is fairly elementary and is omitted. ■

4 Relativistic Energy Bounds

Theorem 1.5: Sketch of Proof.

An upper bound for E_0 is easy to obtain, but it is indirect. Note that

$$|p + \sqrt{\alpha}A(x)| \leq \varepsilon[p + \sqrt{\alpha}A(x)]^2 + (4\varepsilon)^{-1} \quad (4.1)$$

for any $\varepsilon > 0$. Take $\Psi = f(x) \otimes |0\rangle$ with $|0\rangle$ being the Fock space vacuum. Using (4.1)

$$\begin{aligned} (\Psi, H\Psi) &\leq \varepsilon \int_{\mathbb{R}^3} \{\alpha \langle 0|A(x)^2|0\rangle |f(x)|^2 + |\nabla f(x)|^2\} dx + \varepsilon^{-1} \\ &= \frac{\varepsilon \alpha \Lambda^2}{4\pi} + \int |\nabla f|^2 + \frac{1}{4\varepsilon}, \end{aligned} \quad (4.2)$$

since $\langle 0|A(x)^2|0\rangle = (2\pi)^{-3} \int_{|k|<\Lambda} |k|^{-1} dk = \Lambda^2/4\pi^2$. We can now let $f(x) \rightarrow V^{-\frac{1}{2}}$ and take $\varepsilon = (\pi/\alpha)^{1/2} \Lambda^{-1}$, whence

$$E_0 \leq (\alpha/4\pi)^{1/2} \Lambda. \quad (4.3)$$

Now we turn to the lower bound for H .

Step 1: Since $x \rightarrow \sqrt{x}$ is operator monotone,

$$T > T_1 = |p_1 + \sqrt{\alpha}A_1(x)|, \quad (4.4)$$

where the subscript 1 denotes the 1 component (i.e., the x -component) of a vector. By replacing T by T_1 , we are now in a position to remove A_1 by a gauge transformation - but it has to be an *operator-valued gauge transformation*. The use of such a gauge transformation is a novelty, as far as we are aware, in QED.

To effect the gauge transformation, set

$$\varphi(x) = \frac{1}{\sqrt{2V}} \sum_{h,\lambda} \frac{\varepsilon^\lambda(k)_1}{\sqrt{(k)}} [a_\lambda(k) + a_\lambda^*(-k)] \frac{e^{ik_1x_1} - 1}{ik_1} e^{ik_\perp x_\perp} \quad (4.5)$$

with $x_\perp = (x_2, x_3)$. Then $[A_j(x), \varphi(x)] = 0$, $j = 1, 2, 3$ and $-i \frac{\partial}{\partial x_1} \exp[i\varphi(x)] = -A_1(x)$. The unitary $U = \exp[i\varphi(x)]$ is a gauge transformation, but it is *operator-valued*. We have

$$\begin{aligned} U^{-1}|p_1 + A_1(x)|U &= |p_1| \\ U^{-1}a_\lambda(k)U &= a_\lambda(k) + f_\lambda(k, x) \\ U^{-1}a_\lambda^*(k)U &= a_\lambda^*(k) + \bar{f}_\lambda(k, x) \\ \tilde{H}_f = U^{-1}H_fU &= \sum_{k,\lambda} |k| [a_\lambda^*(k) + \bar{f}_\lambda(k, x)] [a_\lambda(k) + f_\lambda(k, x)] \end{aligned} \quad (4.6)$$

with

$$f_\lambda(k, x) = \sqrt{\frac{\alpha}{2V}} \sum_{k,\lambda} \frac{\varepsilon_\lambda(k)_1}{|k|} \frac{e^{-ik_1x_1} - 1}{k_1} e^{-ik_\perp x_\perp}. \quad (4.7)$$

Since $p_\perp$ does not appear in our new Hamiltonian,

$$\tilde{H} = U^{-1} H U = |p_1| + \tilde{H}_f, \quad (4.8)$$

the variable $x_\perp$ appears only as a parameter, and thus we can set $x_\perp = \text{constant} = (0,0)$, by translation invariance, and replace $\mathbb{R}^3$ by $\mathbb{R}^1 = \mathbb{R}$.

From now on $x_1 = x$ and, $p_1 = p = -i d/dx$.

Step 2: The dependence on x now appears in $\tilde{H}_f$ instead of in the kinetic energy, $|p|$. For each x we can try to put $\tilde{H}_f$ into its ground state, which is that of a displaced harmonic oscillator. But, since this state depends on x , to do so will require a great deal of kinetic energy, $|p_1|$.

Let Ψ be a normalized wave-function, i.e., a function on $L^2(\mathbb{R}) \otimes \mathcal{F}$. We write it as ψ_x where $\psi_x \in \mathcal{F}$. Thus, with $\langle \cdot, \cdot \rangle$ denoting the inner product on $\mathcal{F}$, $\int_{\mathbb{R}} \langle \psi_x, \psi_x \rangle dx = 1$.

Decompose $\mathbb{R}$ as the disjoint union of intervals of length ℓ/Λ , where ℓ is a parameter to be determined later. Denote these intervals by I_j , $j = 1, 2, \dots$. A simple Poincaré type inequality gives, for $g : L^2(\mathbb{R}) \rightarrow \mathbb{C}$,

$$(g, |p|g) \geq C_1 \frac{\Lambda}{\ell} \sum_j \int_{I_j} \{ |g(x)|^2 - |\bar{g}_j|^2 \} dx,$$

where $\bar{g}_j = \frac{\Lambda}{\ell} \int_{I_j} g(x) dx$ is the average of g in I_j . Then

$$(\Psi, |p|\Psi) \geq C_1 \frac{\Lambda}{\ell} \sum_j \int_{I_j} \{ \langle \psi_x, \psi_x \rangle - \langle \bar{\psi}_j, \bar{\psi}_j \rangle \} dx. \quad (4.9)$$

Step 3: Next, we analyze $\tilde{H}_f$. We think of this as an operator on $\mathcal{F}$, parameterized by $x \in \mathbb{R}$. We would like $\tilde{H}_f$ to have a gap so we define

$$H_x = \frac{\Lambda}{2} \sum_{\varepsilon \Lambda \leq |k| \leq \Lambda} \sum_{\lambda} [a_{\lambda}^+(k) + \bar{f}_{\lambda}(k, x)] \cdot [h.c.] \quad (4.10)$$

Clearly, $\tilde{H}_f \geq H_x$ and

$$(\Psi, \tilde{H} \Psi) \geq \frac{\Lambda}{\ell} \sum_j \int_{I_j} \langle \psi_x, \psi_x \rangle - \langle \bar{\psi}_j, \bar{\psi}_j \rangle + \langle \psi_x, H_x \psi_x \rangle dx. \quad (4.11)$$

For each interval I_j we can minimize (4.11) subject to $\int_{I_j} \langle \psi_x, \psi_x \rangle dx$ fixed. This leads to

$$(h_j \psi)_x = e \psi_x \quad (4.12)$$

with

$$(h_j \psi)_x = \frac{\Lambda}{\ell} \psi_x - \frac{\Lambda}{\ell} \bar{\psi}_j + H_x \psi_x \quad (4.13)$$

Obviously, this eigenvalue problem (4.12, 4.13) is the same for all intervals I_j , so we shall drop the subscript j and try to find the minimum e .

A lower bound to h_j (and hence to e) can be found by replacing H_x by

$$\hat{H}_x = \frac{\Lambda}{2}(1 - \Pi_x) ,$$

where $\Pi_x = |g_x\rangle\langle g_x|$ is the projector onto the ground state, $|g_x\rangle$, of H_x .

If we substitute $\hat{H}_x$ into (4.13) the corresponding eigenvalue equation (4.12) becomes soluble. Multiply (4.12) on the left by $\langle g_x|$, whence

$$\left(\frac{\Lambda}{\ell} - e\right) \langle g_x, \psi_x \rangle = \frac{\Lambda}{\ell} \langle g_x, \bar{\psi} \rangle \quad (4.14)$$

Then, substitute (4.14) into (4.13) and integrate $\int_I dx$ to find

$$\frac{1}{2} \Lambda^3 \ell^{-2} \left(\int \Pi_x dx \right) \bar{\psi} = \left(\frac{\Lambda}{\ell} - e \right) \left(\frac{\Lambda}{2} - e \right) \bar{\psi} . \quad (4.15)$$

We know that $e < \Lambda/2$ because we could take $\psi_x = \text{constant}$ as a trial function, and then use $\tilde{H}_x \leq \Lambda/2$. Also, $e < \Lambda/\ell$, because we could take $\Psi = \delta_{x_0} |g_{x_0}\rangle$.

Step 4: Eq. (4.15) will give us a lower bound to e if we can find an upper bound to $Y = (\Lambda/\ell) \int_I \Pi_x dx$. To do this note that

$$\begin{aligned} Y^2 &\leq \text{Trace } Y^2 = \left(\frac{\Lambda}{\ell}\right)^2 \int_I \int_I |\langle g_x, g_y \rangle|^2 dx dy \\ &= \left(\frac{\Lambda}{\ell}\right)^2 \int_I \int_I \exp\left\{-\frac{\alpha}{2V} \sum_{\varepsilon\Lambda \leq |k| \leq \Lambda} \sum_{\lambda} |f_{\lambda}(k, x) - f_{\lambda}(k, y)|^2 dx dy\right\} \quad (4.16) \end{aligned}$$

Noting that $\sum_{\lambda=1}^2 e_{\lambda}(k)_1^2 = 1 - k_1^2/k^2$, the quantity $\{ \quad \}$ in (4.16) becomes (as $V \rightarrow \infty$)

$$\{ \quad \} = -\frac{2\alpha}{(2\pi)^3} \int_{\Lambda/2 < |k| < \Lambda} \frac{1}{|k|^3 k_1^2} (k_{\perp}^2) \left[\sin \frac{k_1}{2} (x - y) \right] \quad (4.17)$$

After some algebra we find that

$$\left(\frac{1}{\ell} - \frac{e}{\Lambda}\right) \left(\frac{1}{2} - \frac{e}{\Lambda}\right) \leq \frac{1}{2\ell} (\text{Trace } Y^2)^{1/2} \leq \frac{1}{2\ell} \sqrt{K_{\ell}(\alpha)} \quad (4.18)$$

where

$$\begin{aligned} K_{\ell}(\alpha) &= \int_0^1 \int_0^1 \exp \left[-\alpha \frac{\ell}{\pi^2} |x - y| \int_0^{|x-y|\ell/4} \left(\frac{\sin t}{t} \right)^2 dt \right] dx dy . \\ &\leq \int_{-1/2}^{1/2} \exp[-\alpha x^2 \ell^2 / 8\pi] dx . \quad (4.19) \end{aligned}$$

We find that

$$\begin{aligned} K_\ell(\alpha) &\sim 1 - \alpha\ell^2/96\pi, & \ell^2\alpha \text{ small} \\ &\sim \sqrt{2}\pi(\alpha\ell^2)^{1/2}, & \ell^2\alpha \text{ large} \end{aligned} \quad (4.20)$$

If α is small we take $\ell \sim \alpha^{-1/2}$. If α is large we take $\ell = 2$. This establishes our theorem for $N = 1$. ■

Theorem 1.6: Sketch of Proof.

For $N > 1$ we can decompose $\mathbb{R}^3$ into cubic boxes $B_j, j = 1, 2, 3, \dots$ of size $\ell\Lambda$ and “borrow” $\frac{1}{2}|p + A(x)|^2$ kinetic energy from each particle. That is, $H_N = H_N^{1/2} + \frac{1}{2}T_N$ with $T_N = \sum_{i=1}^N T(x_i)$. The Pauli principle will then yield an energy for $\frac{1}{2}T_N$ that is bounded below by (const.) $(n_{j-1})^{4/3}$, where n_j is the particle number in box B_j .

References

- [BFG] L. Bugliaro, J. Fröhlich and G.M. Graf, *Stability of quantum electrodynamics with nonrelativistic matter*, Phys.Rev. Lett. **77** (1996), 3494–3497.
- [BFS] V. Bach, J. Fröhlich and I.M. Sigal, *Renormalization group analysis of spectral problems in quantum field theory*, Adv. Math. **137** (1998), 205–298; *Quantum electrodynamics of confined nonrelativistic particles*, *ibid* 299–395.
- [F] C. Fefferman, *Stability of Coulomb systems in a magnetic field*, Proc. Natl. Acad. Sci. USA, **92** (1995), 5006–5007.
- [FFG] C. Fefferman, J. Fröhlich and G.M. Graf, *Stability of nonrelativistic quantum mechanical matter coupled to the (ultraviolet cutoff) radiation field*, Proc. Natl. Acad. Sci. USA **93** (1996), 15009–15011; *Stability of ultraviolet cutoff quantum electrodynamics with non-relativistic matter*, Commun. Math. Phys. **190** (1997), 309–330.
- [H] H. Hogreve, Math. Reviews **99e:81051a**, **b** Amer. Math. Soc. (1999).
- [LL] E.H. Lieb and M. Loss, *Remarks about the ultraviolet problem in QED*, (in preparation).
- [LLS] E.H. Lieb, M. Loss and J. P. Solovej, *Stability of Matter in Magnetic Fields*, Phys. Rev. Lett. **75**, 985–989 (1995).
- [LSS] E.H. Lieb, H. Siedentop and J.P. Solovej, *Stability and Instability of Relativistic Electrons in Magnetic Fields*, J. Stat. Phys. **89** (1997), 37–59.
- [LY] E.H. Lieb and K. Yamazaki, *Ground State Energy and Effective Mass of the Polaron*, Phys. Rev. **111** (1958), 728–733.
- [OY] T. Okamoto and K. Yajima, *Complex scaling technique in nonrelativistic massive QED*, Ann. Inst. H. Poincaré Phys. Théor. **42** (1985), 311–327.

Ground states in non-relativistic quantum electrodynamics

Marcel Griesemer^{1,*}, Elliott H. Lieb^{2,**}, Michael Loss^{3,***}

¹ Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294, USA (e-mail: marcel@math.uab.edu)

² Departments of Physics and Mathematics, Jadwin Hall, Princeton University, P. O. Box 708, Princeton, NJ 08544, USA (e-mail: lieb@princeton.edu)

³ School of Mathematics, Georgia Tech, Atlanta, GA 30332, USA
(e-mail: loss@math.gatech.edu)

Oblatum 21-IX-2000 & 8-IV-2001

Published online: 18 June 2001

Abstract. The excited states of a charged particle interacting with the quantized electromagnetic field and an external potential all decay, but such a particle should have a true ground state – one that minimizes the energy and satisfies the Schrödinger equation. We prove quite generally that this state exists for *all values* of the fine-structure constant and the ultraviolet cutoff. We also show the same thing for a many-particle system under physically natural conditions.

1 Introduction

An established picture of an atom or molecule is that even in the presence of a quantized radiation field there is a ground state. The excited states that exist in the absence of coupling to the field are expected to melt into resonances, which means that they eventually decay with time into the ground state plus free photons. This picture has been established by Bach, Fröhlich and Sigal in [8] for sufficiently small values of the various parameters that define the theory. Here we show that a ground state exists for all values of the parameters (including a variable g -factor) in the one particle case and, under a physically appropriate assumption, in the many-particle case.

* Work partially supported by the Faculty Development Program of UAB.

** Work partially supported by U.S. National Science Foundation grant PHY 98-20650-A01.

*** Work partially supported by U.S. National Science Foundation grant DMS 00-70589.

© copyright 2000 by the authors. Reproduction of this article, in its entirety, by any means, is permitted for non-commercial purposes.

We know that the Hamiltonian for the system is bounded below, so a ground state energy always exists in the sense of being the infimum of the spectrum, but the existence of a genuine normalizable solution to the eigenvalue equation is a more delicate matter that has received a great deal of attention, especially in recent years. A physically simple example where no ground state exists (as far as we believe now) is the free particle (i.e., particle plus field). In the presence of an external potential, however, like the Coulomb potential of a nucleus, a ground state should exist.

The difficulty in establishing this ground state comes from the fact that the bottom of the spectrum lies in the continuum (i.e., essential spectrum), not below it, as is the case for the usual Schrödinger equation. We denote the bottom of spectrum of the free-particle Hamiltonian for N particles with appropriate statistics by $E^0(N)$. The “free-particle” Hamiltonian includes the interparticle interaction (e.g., the Coulomb repulsion of electrons) but it does not include the interaction with a fixed external potential, e.g., the interaction with nuclei. When the latter is included we denote the bottom of the spectrum by $E^V(N)$. It is not hard to see in many cases that $E^V(N) < E^0(N)$, but despite this inequality $E^V(N)$ is, nevertheless, the bottom of the essential spectrum. The reason is that we can always add arbitrarily many, arbitrarily ‘soft’ photons that add arbitrarily little energy. It is the soft photon problem that is our primary concern here.

The main point of this paper is to show how to overcome this infrared problem and to show, quite generally for a one-particle system, that a ground state exists for all values of the particle mass, the coupling to the field (fine-structure constant $\alpha = e^2/\hbar c$), the magnetic g -factor, and the ultraviolet cutoff Λ of the electromagnetic field frequencies, provided a bound state exists when the field is turned off. This result implies, in particular, that for a fixed ultraviolet cutoff renormalization of the various physical quantities will not affect the existence of a ground state. Of course, nothing can be said about the limit as the cutoff tends to infinity. We also include a large class of interactions much more general than the usual Coulomb interaction.

The model we discuss has been used quite frequently in field theory. In its classical version it was investigated by Kramers [12] who seems to have been the first to point out the possibility of renormalization. The quantized version was investigated by Pauli and Fierz [24] in connection with scattering theory. Most importantly, it was used by Bethe [9] to obtain a surprisingly good value for the Lamb shift.

Various restricted versions of the problem have been attacked successfully. In the early seventies Fröhlich investigated the infrared problem in translation invariant models of scalar electrons coupled to scalar bosons [13]. In [14] he proved that for an electron coupled to a massive field a unique ground state exists for fixed total momentum. External potentials were not considered in these papers.

The first rigorous result on the bound state problem, to our knowledge, is due to Arai and concerns one particle confined by an x^2 -potential, the interaction with the photons being subject to the dipole-approximation.

For this model, which is explicitly solvable, Arai proved existence and uniqueness of the ground state [2]. Later, Spohn showed by perturbation theory that this result extends to perturbed x^2 -potentials [26]. No bounds on the parameters were needed to obtain these results but the methods obviously do not admit extensions to more realistic models.

Bach, Fröhlich and Sigal [6,7] initiated the study of the full nonrelativistic QED model (the same model as considered in the present paper) under various simplifying assumptions. In [8] the existence of a ground state in this model for particles subject to an external binding potential was proved for $\alpha\Lambda$ small enough. The main achievement of this paper is the elimination of the earlier simplifying assumptions, especially the infrared regularization. This is the first, and up to now the only paper where a ‘first principles’ QED model was successfully analyzed, but with a restricted parameter range. A weaker result for the same system but with simplifications such as infrared regularization were independently obtained by Hiroshima by entirely different methods; he also showed uniqueness of the ground state, assuming its existence [17].

In a parallel development Arai and Hirokawa [3,4], Spohn [28], and Gerard [15] investigated the ground-state problem for systems similar to the one of Bach et al in various degrees of mathematical generality. Arai and Hirokawa proved existence in what they called a “generalized spin-boson model”. If specialized to the case of a non-relativistic N -particle system interacting with bosons, their result proves the existence of a ground state for confined particles and small α . This result was extended in a recent preprint [4] to account for non-confined particles and systems with the true infrared singularity of QED. Concrete results in the infrared-singular case concern only special models, however, such as the Wigner-Weisskopf Hamiltonian which describes a two-level atom. Hirokawa continued this work in a recent preprint [16]. Spohn and Gerard both proved existence of a ground state for a confined particle and arbitrary coupling constant, the result of Gerard being somewhat more general [28,15].

For the existence of a ground state in the case of massive bosons, which is a typical intermediate result in the cited works, a short and elegant proof was given by Dereziński and Gerard [11,15] for the case of linear coupling, in which the A^2 term is omitted, and a confining potential. Some of their ideas are used in our paper.

The Hamiltonian for N -particles has four parts which are described precisely in the next section,

$$H^V = T + V + I + H_f . \quad (1)$$

The dependence of H^V on N is not noted explicitly. The first term, $T = \sum_{j=1}^N T_j$, is the kinetic energy with “minimal coupling” in the Coulomb gauge (i.e., p is replaced by $p + \sqrt{\alpha} A$, where A is the magnetic vector potential satisfying $\operatorname{div} A = 0$), $V(X)$, with $X = (x_1, x_2, \dots, x_N)$ is the external potential, typically a Coulomb attraction to one or more nuclei. In

any case, we assume that V is a sum of one-body potentials, i.e.,

$$V(X) = \sum_{j=1}^N v(x_j) .$$

The particle-particle interaction, I , has the important feature that it is translation invariant and, of course, symmetric in the particle labels. Both I and V could be spin dependent, but we shall not burden the notation with this latter possibility. Typically I is a Coulomb repulsion, but we do not have to assume that I is merely a sum of two-body potentials. The only requirements are: (1) the negative part of v vanishes at infinity; (2) the negative part of I satisfies clustering, i.e., the intercluster part of I_- tends to 0 when the spacing between any two clusters tends to infinity; (3) $V_- + I_-$ are dominated by the kinetic energy as in (5). Another assumption we make (in the N -particle case) is that there is binding, as described below.

The natural choice for the kinetic energy is the “Pauli operator” $T = (p + \sqrt{\alpha} A(x))^2 + \sqrt{\alpha} \sigma \cdot B$, but we can generalize this to include the case of the usual kinetic energy $T = (p + \sqrt{\alpha} A(x))^2$ by introducing a ‘ g -factor’, $g \in \mathbb{R}$. Thus, we take

$$T = (p + \sqrt{\alpha} A(x))^2 + \frac{g}{2} \sqrt{\alpha} \sigma \cdot B . \quad (2)$$

Note that T is a positive operator if $0 \leq g \leq 2$ and may not be otherwise. Nevertheless, the Hamiltonian is always bounded below because of the ultraviolet cutoff we shall impose on the A field, which implies that $(g/2)\sqrt{\alpha} \sigma \cdot B + H_f$ is always bounded below.

We believe that the “relativistic” operator $T = |p + \sqrt{\alpha} A(x)|$, presents no real difficulty either, but we do not want to overburden this paper with a lengthy proof. This problem is currently under investigation.

A model that is frequently discussed is the “Pauli-Fierz” model, but it is not entirely clear how this is defined since several variants appear in [24]. One version uses $T = (p + \sqrt{\alpha} A(x))^2 + \sqrt{\alpha} \sigma \cdot B$, which is one of the models under consideration. Another variant uses a linearized version of this operator, $T = p^2 + 2\sqrt{\alpha} p \cdot A(x) + \sqrt{\alpha} \sigma \cdot B$ or $T = p^2 + 2\sqrt{\alpha} p \cdot A(x)$. These variants are *not gauge invariant* and, therefore, depend on the choice of gauge for A . Our method is applicable to these linearized models in some gauges, but not in others. We omit further discussion of this point since these variants are not the most relevant ones for quantum electrodynamics.

There is one important point that as far as we know, has not been mentioned in the QED context. This is the binding condition. Our proof of the existence of a ground state uses, as input, the assumption that

$$E^V(N) < E^V(N') + E^0(N - N') \quad \text{for all } N' < N. \quad (3)$$

Our work can be generalized (but we shall not do so here) to the case in which the external potential is that of attractive nuclei and these nuclei

are also dynamical particles. Then, of course, it is necessary to work in the center of mass system and then (3) must be replaced by the condition that $E^V(N)$ is less than the lowest two-cluster threshold. While this condition, or (3) in the static case, are physically necessary for the existence of a ground state, the validity of these conditions cannot be taken for granted.

We prove inequality (3) for one particle ($N = 1$) quite generally, using only the assumption that the ordinary Schrödinger operator $p^2 + v$ has a negative energy ground state. This certainly holds for the Coulomb potential. Indeed, one could expect, on physical grounds, that there could be binding even if $p^2 + v$ has no negative energy bound state, because the interaction with the field increases the effective mass of the particle – and hence the binding energy. The same argument shows that when there are N particles at least one of them is necessarily bound, i.e., $E^V(N) < E^0(N)$. When we consider more than one particle, we are not able to show (3) for all $N' > 1$, even if $\sum p_j^2 + I + V$ has a ground state.

In the Coulomb case, it is possible to show (but we shall not do so here) that condition (3) is satisfied if the nuclear charge Z is large enough. The basic idea is that if breakup into two groups, one of them with N' particles close to the nucleus and the second consisting of $N - N'$ particles far away occurs, then there will be an attractive Coulomb tail acting on the separated particles at a distance R away with net attractive potential $(Z - N')/R$. However, to localize one of these particles within a distance R of the nucleus will require a field energy localization error of the order of C/R , by dimensional analysis arguments. If $(Z - N') > C$ then the energy can be decreased by bringing one of the unbound particles close to the nucleus.

Section 2 introduces the precise definition of our problem and the main result Theorem 2.1.

In Sect. 3, we show how to prove that $E^V(1) < E^0(1)$. More generally, $E^V(N) < E^0(N)$ if $v \leq 0$.

Our strategy to establish a ground state is the usual one of showing that a minimizing sequence of trial vectors for the energy actually has a weak limit that, in fact, is a minimizer. The problem here is that one can easily construct minimizing sequences that converge weakly to zero by choosing vectors with too many soft photons. To avoid this we take a special sequence.

To define this sequence we first consider an artificial model in which the photons have a mass, i.e., $\omega(k) = \sqrt{k^2 + m^2}$. Here there is no soft photon problem and we show in Sect. 4 that this model has a ground state Φ_m .

In Sect. 5 we show that as $m \rightarrow 0$ the Φ_m sequence is minimizing. Then in Sect. 6 we use the Schrödinger equation for Φ_m to deduce certain properties of Φ_m which we call infrared bounds. One of these was proved in [8] but we need one more, which is new.

With these bounds we can show in Sect. 7 that Φ_m has a strong limit as $m \rightarrow 0$, which is a minimizer for H^V .

Acknowledgement. We thank Professor Fumio Hiroshima for a useful correspondence concerning equation (53) and for sending us his preprint [18].

2 Definitions and main theorem

The Hamiltonian for N particles interacting with the quantized radiation field and with a given external potential $V(X)$, with $X = (x_1, x_2, \dots, x_N)$ and $x_j \in \mathbb{R}^3$, is

$$H^V = \sum_{j=1}^N \left\{ (p_j + \sqrt{\alpha} A(x_j))^2 + \frac{g}{2} \sqrt{\alpha} \sigma_j \cdot B(x_j) \right\} + V(X) + I(X) + H_f. \tag{4}$$

The unit of energy is $Mc^2/2$, where M is the particle mass, the unit of length is $\ell_c = 2\hbar/Mc$, twice the Compton wavelength of the particle, and $\alpha = e^2/\hbar c$ is the dimensionless “fine structure constant” ($= 1/137$ in nature). The electric charge of the particle is e . The unit of time is the time it takes a light wave to travel a Compton wavelength, i. e., the speed of light is $c = 1$.

The operator $p = -i\nabla$, while A is the (ultraviolet cutoff) magnetic vector potential (we use the Coulomb, or radiation gauge). The unit of $A^2(x)$ is $Mc^2/2\ell_c$. The magnetic field is $B = \text{curl} A$ and the unit of B is $\alpha^{3/2}$ times the quantity $M^2 e^3 c / 4\hbar^3$, which is the value of B for which the magnetic length $(\hbar c / eB)^{1/2}$ equals the Bohr radius $2\hbar^2 / Me^2$.

The reader might wonder why we use these units, which seem to be more appropriate for a relativistic theory than for the nonrelativistic theory we are considering. Why not use the Bohr radius as the unit of length, for example? Our reason is that we want to isolate the electric charge, which is the quantity that defines the interaction of matter with the electromagnetic field, in precisely one place, namely α . “Atomic units” have the charge built into the length, etc. and we find this difficult to disentangle.

The Hilbert space is an appropriate subspace of

$$\mathcal{H} = \otimes^N L^2(\mathbb{R}^3; \mathbb{C}^2) \otimes \mathcal{F},$$

where $\mathcal{F}$ is the Fock space for the photon field. We have in mind Fermi statistics (the antisymmetric subspace of $\otimes^N L^2(\mathbb{R}^3; \mathbb{C}^2)$) and the $\mathbb{C}^2$ is to accomodate the electron spin. We can also deal with “Boltzmann statistics”, in which case we would set $g = 0$ and use $\otimes^N L^2(\mathbb{R}^3)$, or bose statistics, in which case we would set $g = 0$ and use the symmetric subspace of $\otimes^N L^2(\mathbb{R}^3)$. These generalizations are mathematically trivial and we do not discuss them further.

For our purposes we assume that for every $\varepsilon > 0$ there exists a constant $a(\varepsilon)$ such that the negative part of the potentials, $V_-(X)$ and $I_-(X)$, satisfy

$$V_- + I_- \leq \varepsilon \sum_{j=1}^N p_j^2 + a(\varepsilon) \quad (5)$$

as quadratic forms on $\mathcal{H}$.

The vector potential is

$$A(x) = \sum_{\lambda=1,2} \int_{|k| < \Lambda} \frac{1}{\sqrt{|k|}} [\varepsilon_\lambda(k) a_\lambda(k) + \varepsilon_\lambda(-k) a_\lambda^*(-k)] e^{ik \cdot x} d^3k \quad (6)$$

where the operators a_λ, a_λ^* satisfy the usual commutation relations

$$[a_\lambda(k), a_\nu^*(q)] = \delta(k - q) \delta_{\lambda,\nu}, \quad [a_\lambda(k), a_\nu(q)] = 0, \quad \text{etc} \quad (7)$$

and the vectors $\varepsilon_\lambda(k)$ are the two possible orthonormal polarization vectors perpendicular to k . They are chosen for convenience in (59,60).

The number Λ is the ultraviolet cutoff on the wavenumbers k . Our results hold for all finite Λ . The details of the cutoff in (6) are quite unimportant, except for the requirement that rotation symmetry in k -space is maintained. E.g., a gaussian cutoff can be used instead of our sharp cutoff. We avoid unnecessary generalisations.

The field energy H_f , sometimes called $d\Gamma(\omega)$ is given by

$$H_f = \sum_{\lambda=1,2} \int_{\mathbb{R}^3} \omega(k) a_\lambda^*(k) a_\lambda(k) d^3k. \quad (8)$$

The energy of a photon is $\omega(k)$ and the physical value of interest to us is

$$\omega(k) = |k|, \quad (9)$$

in our units. Indeed, any continuous function that is bounded below by $\text{const} \cdot |k|$ for small $|k|$ is acceptable. In the process of proving the existence of a ground state for H we will first study the unphysical ‘‘massive photon’’ case, in which

$$\omega_m(k) = \sqrt{k^2 + m^2} \quad (10)$$

for some $m > 0$, called the ‘photon mass’.

In the remainder of this paper, unless otherwise stated, we shall always assume that there is no restriction on α , Λ and g and that $\omega(k)$ can be either as in (9) or as in (10).

By Lemma A.5 we see easily that H^V is bounded below for all values of the parameters, including $m = 0$. Thus, H^V defines a closable quadratic form and hence a selfadjoint operator, the Friedrich’s extension. We denote this extension again by H^V .

Our main theorem is

2.1. Theorem (Existence of a ground state). *Assume that the binding condition (3) and the condition (5) hold. Then there is a vector Φ in the N -particle Hilbert space $\mathcal{H}$ such that*

$$H^V \Phi = E^V(N) \Phi . \quad (11)$$

3 Upper bound

We shall prove the binding condition (3) for one particle and, with an additional assumption, for the N particle case as well. This is that if N particles are present then at least one of them binds.

As we mentioned before, our requirement that the system without the radiation field has a bound state is somewhat unnaturally restrictive, since one expects that the radiation field *enhances binding*, at least in the single particle case; this has been shown to be true in the “dipole”, or Kramers approximation [19]. We are able to show in the *one-particle* case, that the photon field cannot decrease the binding energy. It is quite possible that there could be binding even when the operator $p^2 + v$ does not have a negative energy state, but we cannot shed any light on that question.

For the one-particle case the situation is less delicate than the N -particle case.

3.1. Theorem (Binding of at least one particle). *Assume that the one-particle Hamiltonian $p^2 + v(x)$ has a negative energy bound state with eigenfunction $\phi(x)$ and energy $-e_0$. Then,*

$$E^V(1) \leq E^0(1) - e_0 , \quad (12)$$

i.e., binding continues to exist when the field is turned on.

For the N -particle case we make the additional assumption that $v(x) \leq 0$ for all x . Then,

$$E^V(N) \leq E^0(N) - e_0 , \quad (13)$$

i.e., at least one particle is bound.

Proof. It suffices to prove that $E^V(N) \leq E^0(N) + \varepsilon - e_0$ for all $\varepsilon > 0$. There is a normalized vector $F \in \mathcal{H}$ such that $(F, H^0 F) < E^0(N) + \varepsilon$. (F is antisymmetric according to the Pauli principle.) We use the notation $\langle \cdot, \cdot \rangle$ to denote the inner product in Fock space and spin space. Then we can write $(F, H^0 F) = \int G(X) d^{3N} X$ with

$$\begin{aligned} G(X) = & \sum_{j=1}^N \left\{ \langle (-i\nabla_j + \sqrt{\alpha} A(x_j)) F, (-i\nabla_j + \sqrt{\alpha} A(x_j)) F \rangle(X) \right. \\ & \left. + \sqrt{\alpha} (g/2) \langle F, \sigma_j \cdot B(x_j) F \rangle(X) \right\} + \langle F, (I + H_f) F \rangle(X) . \end{aligned} \quad (14)$$

As a (unnormalized) variational trial vector we take the vector $\psi = \left[\sum_{j=1}^N \phi(x_j)^2 \right]^{1/2} F$. Recall that $\phi(x) \geq 0$ since ϕ is the ground state of $p^2 + v$. We also recall the Schwarz inequality

$$\left| \frac{\sum_{j=1}^N \phi(x_j) \nabla \phi(x_j)}{\left[\sum_{j=1}^N \phi(x_j)^2 \right]^{1/2}} \right|^2 \leq \sum_{j=1}^N |\nabla \phi(x_j)|^2. \quad (15)$$

Using (15), integration by parts, and the fact that ϕ satisfies the Schrödinger equation $(p^2 + v)\phi = -e_0\phi$, we easily find that

$$\begin{aligned} & (\psi, [H^V - (E^0(N) + \varepsilon - e_0)] \psi) \\ & \leq \int \{G(X) - (E^0(N) + \varepsilon) \langle F, F \rangle(X)\} \sum_{j=1}^N \phi(x_j)^2 d^{3N} X \\ & + \int \sum_{j \neq k} v(x_k) \phi(x_j)^2 \langle F, F \rangle(X) d^{3N} X. \end{aligned} \quad (16)$$

When $N = 1$ the last term in (16) is not present so no assumption about the potential v is needed. When $N > 1$ we can omit the last term because it is negative by assumption.

Now, by the $\mathbb{R}^3$ -translation invariance of H^0 , for every $y \in \mathbb{R}^3$ there is a “translated” vector F_y so that $G(X) \rightarrow G(X + (y, \dots, y))$ and $\langle F_y, F_y \rangle(X) = \langle F, F \rangle(X + (y, \dots, y))$. (This is accomplished by the unitary operator on $\mathcal{H}$ that takes $x_j \rightarrow x_j + y$ for every j and $a_\lambda(k) \rightarrow \exp(ik \cdot y) a_\lambda(k)$.) Thus, if we denote the quantity in $\{ \}$ in (16) by $W(X)$, and if we define ψ_y by replacing F by F_y in the definition of ψ , we have

$$\begin{aligned} \Omega(y) &= (\psi_y, H^V - (E^0 + \varepsilon - e_0) \psi_y) \\ &\leq \int W(X + (y, \dots, y)) \sum_{j=1}^N \phi(x_j)^2 d^{3N} X = \int W(X) \sum_{j=1}^N \phi(x_j - y)^2 d^{3N} X. \end{aligned} \quad (17)$$

Note that $\int \Omega(y) dy \leq N \int W(X) d^{3N} X$. But $\int W(X) d^{3N} X = (F, (H^0 - E^0(N) - \varepsilon) F)$ and this is strictly negative by assumption. Hence, for some $y \in \mathbb{R}^3$ we have that $\Omega(y) < 0$ and thus $\psi_y \neq 0$, which proves the theorem. $\square$

Remark: [Alternative theorem]

It may be useful to note, briefly, a different proof of Theorem 3.1, for *long range* potentials $v(x)$, such as the attractive Coulomb potential $-Z/|x|$, which shows that the bottom of the spectrum of H^V lies strictly below E^0 . Unfortunately, this proof does not show that the difference is at least e_0 . We

sketch it for the one-body case. Using the notation of the proof above, the first step is to replace F by $F_R = u(x_1/R)F$ where u is a smooth function with support in a ball of radius 1. One easily finds that $(F_R, H^0 F_R) / (F_R, F_R) = E^0 + \varepsilon + c/R^2$, where c is a constant that depends only on u and not on ε and R . On the other hand $(F_R, V F_R) / (F_R, F_R) \leq -Z/R$, to use the Coulomb potential as an example. To complete the argument, choose $R = 2c/Z$ and then choose $\varepsilon = c/R^2$. What we have used here is the fact that localization ‘costs’ a kinetic energy R^{-2} , while the potential energy falls off slower than this, e.g., R^{-1} .

4 Ground state with massive photons

As we emphasized in the introduction, not every minimizing sequence converges to the minimizer for our $m = 0$ problem, i.e., with $\omega(k) = |k|$. The situation is much easier for the massive case (10). The Hamiltonian in this case is given by (4) and H_f is given by (8) with (10). To emphasize the dependence on m we denote this Hamiltonian and field energy by H_m^V and $H_f(m)$, respectively. Likewise, $E^V(m, N)$ and $E^0(m, N)$ denote the mass dependent energies, as defined before.

We emphasize that the vector potential is still given by (6), but we could, if we wished, easily replace $|k|^{-1/2}$ in (6) by $(k^2 + m^2)^{-1/4}$.

It will be shown in this section that H_m^V has a ground state. More precisely we prove

4.1. Theorem (Existence of ground state). *Assume that for some fixed value of the ultraviolet cutoff Λ there is binding for the Hamiltonian H_m^V , that is, $E^V(m, N) < \Sigma^V(m, N)$ where $\Sigma^V(m, N) = \min\{E^V(m, N') + E^0(m, N - N') : \text{all } N' < N\}$ is the “lowest two-cluster threshold”. Then $E^V(m, N)$ is an eigenvalue, i.e., there exists a state Φ_m in $\mathcal{H}$ such that $H_m^V \Phi_m = E^V(m, N) \Phi_m$.*

Proof. Let us first show that it suffices to prove that for any normalized sequence Ψ^j , $j = 1, 2, \dots$, (not necessarily minimizing) tending weakly to zero

$$\liminf_{j \rightarrow \infty} (\Psi^j, H_m^V \Psi^j) > E^V(m, N) . \quad (18)$$

To prove this let Φ^j be some *minimizing* sequence, i.e., assume that

$$\|\Phi^j\| = 1 , \quad (19)$$

and that

$$(\Phi^j, H_m^V \Phi^j) \rightarrow E^V(m, N) . \quad (20)$$

By the Banach Alaoglu Theorem we can assume that this sequence, as well as the sequence $H_m^V \Phi^j$ converge weakly in the sense that for any $\Psi \in \mathcal{H}$ with $(\Psi, H_m^V \Psi) < \infty$ we have that

$$(\Psi, H_m^V \Phi^j) \rightarrow (\Psi, H_m^V \Phi_m), \quad (21)$$

where Φ_m is the weak limit of Φ^j . Our goal is to show that $(\Phi_m, H_m^V \Phi_m) = E^V(m, N)$ and that $\|\Phi_m\| = 1$.

Write $\Phi^j = \Phi_m + \Psi^j$. Obviously Ψ^j as well as $H_m^V \Psi^j$ go weakly to zero. Thus

$$\begin{aligned} 0 &= \lim_{j \rightarrow \infty} (\Phi^j, (H_m^V - E^V(m, N)) \Phi^j) \\ &= \lim_{j \rightarrow \infty} ((\Phi_m + \Psi^j), (H_m^V - E^V(m, N))(\Phi_m + \Psi^j)) \\ &= \lim_{j \rightarrow \infty} (\Psi^j, (H_m^V - E^V(m, N)) \Psi^j) + (\Phi_m, (H_m^V - E^V(m, N)) \Phi_m) \end{aligned}$$

where we used that the cross terms vanish. Since $H_m^V - E^V(m, N) \geq 0$ this shows that Φ_m minimizes the energy, and, furthermore, that

$$0 \geq \lim_{j \rightarrow \infty} (\Psi^j, (H_m^V - E^V(m, N)) \Psi^j) \geq \delta \liminf_{j \rightarrow \infty} \|\Psi^j\|^2$$

for some positive constant δ . The second inequality is trivial if $\liminf_{j \rightarrow \infty} \|\Psi^j\|^2 = 0$ and otherwise follows from our assumption (18). This proves that Ψ^j converges strongly to zero along a subsequence, which implies that $\|\Phi_m\| = 1$. Hence Φ_m is a normalized ground state. Thus, it suffices to prove (18).

The steps that lead to a proof of (18) are quite standard. The only difficulty is that one has to localize in Fock space, which we describe first. We follow [11] with some necessary modifications and some simplifications.

Recall that, when the $a_\lambda^\#$ operators are viewed in x -space

$$a_\lambda(f) : \mathcal{F} \rightarrow \mathcal{F}, \quad a_\lambda^*(g) : \mathcal{F} \rightarrow \mathcal{F}, \quad (22)$$

they obey the commutation relations

$$[a_\lambda(f), a_\lambda^*(g)] = \int_{\mathbb{R}^3} \bar{f}(x) g(x) d^3x =: (f, g). \quad (23)$$

Consider now two smooth localization functions j_1 and j_2 that satisfy $j_1^2 + j_2^2 = 1$ and j_1 is supported in a ball of radius P . The first derivatives of j_1 and j_2 are of order $1/P$.

The operators

$$\begin{aligned} c_\lambda(f) &= a_\lambda(j_1 f) \otimes \mathcal{I} + \mathcal{I} \otimes a_\lambda(j_2 f), \\ c_\lambda^*(g) &= a_\lambda^*(j_1 g) \otimes \mathcal{I} + \mathcal{I} \otimes a_\lambda^*(j_2 g) \end{aligned} \quad (24)$$

act both on the space $\mathcal{F} \otimes \mathcal{F}$. Note that

$$[c_\lambda(f), c_\lambda^*(g)] = (f, g) . \quad (25)$$

Thus, these new creation and annihilation operators create another Fock space $\mathcal{F}^l$ that is a subspace of $\mathcal{F} \otimes \mathcal{F}$ and is isomorphic to the old Fock space $\mathcal{F}$. Hence, there exists a map

$$U : \mathcal{F} \rightarrow \mathcal{F}^l \quad (26)$$

that is an invertible isometry between Fock spaces. It is uniquely specified by the properties

$$a^\# = U^* c^\# U , \quad (27)$$

and the vacuum in $\mathcal{F}$ is mapped to the vacuum in $\mathcal{F} \otimes \mathcal{F}$.

The map U^* is defined on $\mathcal{F}^l$ only, but we can extend it to all of $\mathcal{F} \otimes \mathcal{F}$ by setting $U^*F = 0$ whenever $F \in \mathcal{F} \otimes \mathcal{F}$ is perpendicular to $\mathcal{F}^l$. In other words U^* is a partial isometry between Fock spaces where $U^*U = \mathcal{I}$ on $\mathcal{F}$, and where UU^* is the orthogonal projection onto $\mathcal{F}^l$. We continue to denote the extended map by U^* .

Let ϕ and $\bar{\phi}$ be smooth nonnegative functions, with $\phi^2 + \bar{\phi}^2 = 1$, ϕ identically one on the unit ball, and vanishing outside the ball of radius 2. Set $\phi_R(X) = \phi(X/R)$. It is a standard calculation to show that for any Ψ with finite energy

$$\begin{aligned} (\Psi, H_m^V \Psi) &= (\phi_R \Psi, H_m^V \phi_R \Psi) + (\bar{\phi}_R \Psi, H_m^V \bar{\phi}_R \Psi) \\ &\quad - (\Psi, (\nabla \phi_R)^2 \Psi) - (\Psi, (\nabla \bar{\phi}_R)^2 \Psi) . \end{aligned} \quad (28)$$

The last two terms in (28) are bounded by $const./R^2$.

One goal will be to show that for any Ψ with finite energy

$$\begin{aligned} (\Psi, \phi_R H_m^V \phi_R \Psi) &= \\ (\Psi, \phi_R U^* \{ H_m^V \otimes \mathcal{I} + \mathcal{I} \otimes H_f \} U \phi_R \Psi) &+ o(1) . \end{aligned} \quad (29)$$

The error term $o(1)$ vanishes as both R and P go to infinity and depends otherwise only on the energy of Ψ . Notice that the invertible map U depends on the cutoff parameter P as well. (29) will be proved in Lemma A.1 in the appendix. The intuition behind the estimate (29) is that localized electrons interact only weakly with far away photons. Those photons are described solely by their own field energy.

An immediate consequence of (29) is the estimate

$$\begin{aligned} (\Psi, \phi_R H_m^V \phi_R \Psi) &\geq (E^V(m, N) + m) \|\phi_R \Psi\|^2 \\ &\quad - m(\phi_R \Psi, U^* \mathcal{I} \otimes P_2 U \phi_R \Psi) + o(1) , \end{aligned} \quad (30)$$

which is obtained by noting that the field energy in the second factor can be estimated from below by

$$H_f \geq m \mathcal{I} - m P_2, \quad (31)$$

where P_2 is the projection onto the vacuum of the second factor of $\mathcal{F} \otimes \mathcal{F}$.

In a further step we prove in Lemma A.3 that the sequence

$$(\phi_R \Psi^j, U^* \mathcal{I} \otimes P_2 U \phi_R \Psi^j) \rightarrow 0 \quad (32)$$

as $j \rightarrow \infty$.

Returning to (28), using Corollary A.2 we have that

$$(\overline{\phi_R} \Psi, H_m^V \overline{\phi_R} \Psi) \geq \Sigma^V(m, N) \|\overline{\phi_R} \Psi\|^2 - o(1), \quad (33)$$

with $o(1)$ going to zero as $R \rightarrow \infty$. Roughly speaking $\overline{\phi}$ forces some of the particles to be far away from the origin. Any such particle configuration can be described by two clusters with no interaction between them. In particular the interaction between these clusters via the radiation field is turned off. This means that each cluster carries its own field energy. To prove this the localization in Fock space is used. Moreover the cluster that is far away from the origin does not interact with the external potential although the repulsion among its particles is still present.

To summarize, by combining (30), (32) and (33) we have proved that

$$\liminf (\Psi^j, H_m^V \Psi^j) \geq (E^V(m, N) + \delta) + o(1) \quad (34)$$

where

$$\delta = \min\{m, \Sigma^V(m, N) - E^V(m, N)\}, \quad (35)$$

and $o(1)$ tends to zero as $R \rightarrow \infty$ and $P \rightarrow \infty$. $\square$

5 A minimizing sequence

We consider the Hamiltonian H_m^V defined in (4) with field energy $H_f(m)$ defined using $\omega_m(k) = \sqrt{k^2 + m^2}$. Our main goal here is Theorem 5.3, which shows that the ground states of the $m > 0$ problem form a minimizing sequence for the $m = 0$ problem.

5.1. Theorem ($E^V(m, N)$ converges to $E^V(0, N)$). As $m \rightarrow 0$,

$$E^V(m, N) \rightarrow E^V(0, N) \quad \text{and} \quad E^0(m, N) \rightarrow E^0(0, N) \quad (36)$$

Proof. First, note that $H_m^V > H_{m'}^V > H_0^V$ if $m > m' > 0$, because ω_m has this same monotonicity property. Therefore, for any sequence of $m \rightarrow 0$, $E^V(m, N)$ is monotonically decreasing and has a sequence-independent, finite limit, which we call E^* , and we note that $E^* \geq E^V(0, N)$. To prove the

opposite, namely $E^* \leq E^V(0, N)$, we shall prove that $E^* \leq E^V(0, N) + 2\varepsilon$ for every $\varepsilon > 0$.

Let $\Phi \in \mathcal{H}$ be normalized and such that $(\Phi, H_0^V \Phi) < E^V(0, N) + \varepsilon$.

We note that $H_m^V < H_0^V + m\mathcal{N}$, where $\mathcal{N}$ is the number operator

$$\mathcal{N} = \sum_{\lambda=1,2} \int_{|k|<\Lambda} a_\lambda^*(k) a_\lambda(k) d^3k. \quad (37)$$

Thus, if we use Φ as a variational function for H_m^V we have $E^V(m, N) < E^V(0, N) + \varepsilon + m(\Phi, \mathcal{N} \Phi)$, and our goal is accomplished *provided that* Φ can be chosen so that $(\Phi, \mathcal{N} \Phi) < \infty$, in addition to $(\Phi, H_0^V \Phi) < E^V(0, N) + \varepsilon$. If a way can be found to modify Φ to another vector $\tilde{\Phi}$ so that $(\tilde{\Phi}, \mathcal{N} \tilde{\Phi}) < \infty$, in addition to $(\tilde{\Phi}, H_0^V \tilde{\Phi}) < E^V(0, N) + 2\varepsilon$ the proof will be complete.

A suitable choice is $\tilde{\Phi} = \Pi_n \Phi$, with Π_n being the projector onto the subspace of $\mathcal{H}$, with n or fewer photons, i.e., $\mathcal{H}_n = \otimes^N L^2(\mathbb{R}^3; \mathbb{C}^2) \otimes \mathcal{F}_{\leq n}$, for an appropriately large n . It is easy to see, with the help of Lemma A.5 that $(\Pi_n \Phi, [V + I] \Pi_n \Phi) \rightarrow (\Phi, [V + I] \Phi)$, that $(\Pi_n \Phi, H_f \Pi_n \Phi) \rightarrow (\Phi, H_f \Phi)$, and $(\Pi_n \Phi, p_j^2 \Pi_n \Phi) \rightarrow (\Phi, p_j^2 \Phi)$ as $n \rightarrow \infty$. The following Lemma 5.2, shows that the other terms converge as well. The same proof works for $E^0(m, N)$. $\square$

5.2. Lemma (Finite photon number approximation). *Let $\Phi \in \mathcal{H}$ be such that $(\Phi, H_f \Phi) < \infty$. Denote by Π_n the projection in $\mathcal{H}$ onto states with photon number less than or equal to n , i.e., Π_n is the projection onto the subspace $\mathcal{H}_n = \otimes^N L^2(\mathbb{R}^3; \mathbb{C}^2) \otimes \mathcal{F}_{\leq n} \subset \mathcal{H}$. Let $\Phi_n = \Pi_n \Phi$. Then we have the following strong convergence as $n \rightarrow \infty$ (in addition to $\Phi_n \rightarrow \Phi$)*

$$\begin{aligned} A(x_j) \Phi_n &\rightarrow A(x_j) \Phi \\ B(x_j) \Phi_n &\rightarrow B(x_j) \Phi. \end{aligned} \quad (38)$$

Proof. Write $A = D + D^*$ where D contains the annihilation operators and D^* contains the creation operators. We omit (x_j) and we omit the vector index of A for simplicity. Since $(\Phi, H_f \Phi) < \infty$ we learn from Lemma A.4 that $D\Phi$ and $D^*\Phi$ are in $\mathcal{H}$. Since $\Pi_{n-1} \rightarrow \mathcal{I}$ strongly,

$$D\Phi_n = \Pi_{n-1} D\Phi \rightarrow D\Phi.$$

The same holds if D is replaced by D^* and $n - 1$ by $n + 1$. This proves the statement for A . The statement for B is proved in the same way. $\square$

The following are two corollaries of Theorem 5.1. The first is the input for Sect. 6. The second is important for showing that it is only necessary to state the “binding” condition (3) once (for $m = 0$). It is a trivial consequence of Theorem 5.1.

5.3. Theorem (Minimizing sequence). *Suppose that $m_1 > m_2 > \dots > 0$ is a sequence tending to zero and suppose that Φ_j , for $j = 1, 2, \dots$ is an approximate minimizer for $H_{m_j}^V$ in the sense that*

$$\delta_j \equiv (\Phi_j, H_{m_j}^V \Phi_j) - E^V(m_j, N) \rightarrow 0 \quad \text{as } j \rightarrow \infty. \quad (39)$$

Then $\Phi_1, \Phi_2, \dots$ is a minimizing sequence for H_0^V .

Proof. $E^V(m_j, N) + \delta_j = (\Phi_j, H_{m_j}^V \Phi_j) > (\Phi_j, H_0^V \Phi_j) \geq E^V(0, N)$. $\square$

5.4. Lemma (Binding without mass implies uniform binding with mass).

Assume (3), i.e., assume that

$$E^V(0, N) < \min\{E^V(0, N') + E^0(0, N - N') : N' < N\} - 2\varepsilon.$$

Then, for all sufficiently small m ,

$$E^V(m, N) < \min\{E^V(m, N') + E^0(m, N - N') : N' < N\} - \varepsilon.$$

Proof. Each of the various energies converges as $m \rightarrow 0$ by Theorem 5.1. $\square$

6 Two infrared bounds

We have seen that any sequence of (approximate) minimizers for the H_m^V problem ($m > 0$) is a minimizing sequence for the H_0^V problem. A crucial point was the possibility of finding an approximate minimizer for the H_0^V problem that has a finite expectation value for the total photon number operator $\mathcal{N}$.

We also know by Corollary 5.4 and Theorem 4.1 that the H_m^V problem has a ground state, Φ_m , and in this section we shall prove two theorems about the soft photon behavior of Φ_m . The first, Theorem 6.1, is about the photon number $a_\lambda^*(k)a_\lambda(k)$ and is based on a method of [8]. The second, Theorem 6.3, which has no antecedent we are aware of, is about the derivative of $a_\lambda^*(k)a_\lambda(k)$ with respect to k .

6.1. Theorem (Photon number bound). *Assume that there is binding, i.e., $E^V(m, N) < \Sigma^V(m, N)$. Assume that Φ_m is a normalized ground state for the many-body Hamiltonian H_m^V , $m \geq 0$. Then*

$$(\Phi_m, a_\lambda^*(k)a_\lambda(k)\Phi_m) < \frac{P\alpha}{|k|}(1 + g^2)\chi_\Lambda(k),$$

where P is a finite constant independent of Φ_m , g , α , and depends on m only via the binding energy $\Sigma^V(m, N) - E^V(m, N) > 0$ and of course on Λ . The function $\chi_\Lambda(k)$ is the characteristic function of the ball of radius Λ .

Remark: We have proved in Corollary 5.4 that the binding energy is a uniformly positive function of m for small m , if the binding energy at $m = 0$ is not zero. Therefore, Theorem 6.1 implies that the expectation value $(\Phi_m, \mathcal{N} \Phi_m)$ of the number operator (37) in the ground state Φ_m is uniformly bounded for small m .

The proof of Theorem 6.1 will be based on the following lemma about exponential decay of eigenfunctions, concerning which there is a vast literature (see [23, 10, 25, 1]). We do not strive at all to get the best exponential decay constants. For us the only really relevant goal is a decay estimate that depends *only* on $\Sigma^V(m, N) - E^V(m, N) > 0$, but not otherwise on m .

6.2. Lemma (Exponential decay). *Let H_m^V be the N -body Hamiltonian in (4) and let Φ_m be a groundstate wave function, which necessarily satisfies the Schrödinger equation*

$$H_m^V \Phi_m = E^V(m, N) \Phi_m . \quad (40)$$

We assume that $\Sigma^V(m, N) - E^V(m, N) > 0$ and choose $\beta > 0$ with $\beta^2 < \Sigma^V(m, N) - E^V(m, N)$. Then

$$\| \exp(\beta|X|) \Phi_m \|^2 \leq C \left(1 + \frac{1}{\Sigma^V(m, N) - E^V(m, N) - \beta^2} \right) \| \Phi_m \|^2 \quad (41)$$

where the constant C does not depend on m .

The strategy of the following proof is probably due to Agmon [1]. We learned it from [20].

Proof. Let $G(X)$ any smooth, bounded function on $\mathbb{R}^{3N}$ with bounded first derivative. We easily compute

$$[[H_m^V - E^V(m, N), G], G] = -2|\nabla G|^2 . \quad (42)$$

We use the Schrödinger equation (40) to compute

$$\begin{aligned} (G \Phi_m, [H_m^V - E^V(m, N)] G \Phi_m) &= -\frac{1}{2} (\Phi_m, [[H_m^V - E(m), G], G] \Phi_m) \\ &= (\Phi_m, |\nabla G|^2 \Phi_m) . \end{aligned} \quad (43)$$

Now we choose G to be

$$\begin{aligned} G(X) &= \chi(X/R) \exp[f(X)], \quad \text{where} \\ f(X) &= \left[\frac{\beta|X|}{1 + \varepsilon|X|} \right], \end{aligned} \quad (44)$$

and where $0 \leq \chi \leq 1$ is a smooth cutoff function that is identically equal to 1 outside the ball of radius 2, and identically zero inside the ball of radius 1. We let $\varepsilon \rightarrow 0$ at the end.

Next, we calculate

$$|\nabla G|^2 = |\nabla \chi|^2 e^{2f} + 2\nabla \chi \cdot \nabla f e^f G + |\nabla f|^2 G^2,$$

and note that the first and second terms are compactly supported in $\mathbb{R}^{3N}$ and each is bounded by a constant C that depends on β and R .

Returning to (43), we obtain, after rearranging terms,

$$(G\Phi_m, (H_m^V - E^V(m, N) - |\nabla f|^2) G\Phi_m) \leq C\|\Phi_m\|^2. \quad (45)$$

Since $|\nabla f| \leq \beta$ we know by Corollary A.2 that

$$\begin{aligned} & \chi (H_m^V - E^V(m, N) - |\nabla f|^2) \chi \\ & \geq (\Sigma^V(m, N) - E^V(m, N) - \beta^2 - o(1)) \chi^2 \quad (R \rightarrow \infty). \end{aligned} \quad (46)$$

In conjunction with (45) this shows that

$$\|G\Phi_m\|^2 \leq \frac{2C}{\Sigma^V(m, N) - E^V(m, N) - \beta^2} \|\Phi_m\|^2$$

for R large enough. After letting $\varepsilon \rightarrow 0$ by monotone convergence a similar bound with G replaced by $\chi \exp(\beta|X|)$ is obtained.

Proof of Theorem 6.1. This proof is a slight modification of the one in [8]. The basic idea is to show that there is effectively no interaction between *localized* particles and low momentum photons. To make this idea explicit we write our Hamiltonian in a gauge different from the usual Coulomb gauge.

To be precise, define

$$\tilde{A}(x) = A(x) - A(0), \quad (47)$$

which is well defined owing to the ultraviolet cutoff. The unitary operator that accomplishes this is $U = \exp[i \sum_{j=1}^N \sqrt{\alpha} x_j \cdot A(0)]$. This is an “operator-valued gauge transformation”. It commutes with $A(x)$ for all x , but not with $a_\lambda(k)$ or with H_f .

Define

$$b_\lambda(k, X) = U a_\lambda(k) U^* = a_\lambda(k) - i w_\lambda(k, X), \quad (48)$$

with $w_\lambda(k, X) = \chi_\Lambda(k) |k|^{-1/2} \varepsilon_\lambda(k) \cdot \sum_{j=1}^N x_j$. The transformed Hamiltonian $\tilde{H}_m$ is

$$\begin{aligned} \tilde{H}_m &= U H_m^V U^* = \sum_{j=1}^N \left\{ (p_j + \sqrt{\alpha} \tilde{A}(x_j))^2 + \frac{g}{2} \sqrt{\alpha} \sigma_j \cdot B(x_j) \right\} \\ &\quad + V + \tilde{H}_f(m) \end{aligned} \quad (49)$$

$$\tilde{H}_f(m) = \sum_{\lambda=1,2} \int_{\mathbb{R}^3} \omega_m(k) b_\lambda^*(k, X) b_\lambda(k, X) d^3 k. \quad (50)$$

To estimate $\|a_\lambda(k)\Phi_m\|$, write

$$a_\lambda(k)\Phi_m = U^*a_\lambda(k)\tilde{\Phi}_m - iw_\lambda(k, X)\Phi_m \quad (51)$$

where $\tilde{\Phi}_m = U\Phi_m$, and note that

$$\|w_\lambda(k, X)\Phi_m\| \leq \sqrt{N} \frac{\chi_\Lambda(k)}{|k|^{1/2}} \| |X| \Phi_m \| . \quad (52)$$

It remains to estimate $\|a_\lambda(k)\tilde{\Phi}_m\|$. By the Schrödinger equation for $\tilde{\Phi}_m$

$$\begin{aligned} (\tilde{H}_m - E^V(m, N))a_\lambda(k)\tilde{\Phi}_m &= [\tilde{H}_m, a_\lambda(k)]\tilde{\Phi}_m \\ &= 2\sqrt{\alpha}|k|^{-1/2}\varepsilon_\lambda(k) \cdot \sum_{j=1}^N (p_j + \sqrt{\alpha}\tilde{A}(x_j))(1 - e^{-ik \cdot x_j})\tilde{\Phi}_m \\ &\quad + i\frac{g}{2}\sqrt{\alpha} \frac{k \wedge \varepsilon_\lambda(k)}{\sqrt{|k|}} \cdot \sum_{j=1}^N \sigma_j e^{-ik \cdot x_j} \tilde{\Phi}_m \\ &\quad - \omega_m(k)b_\lambda(k, X)\tilde{\Phi}_m . \end{aligned} \quad (53)$$

While this equation is correct, the derivation is somewhat formal. This is rigorously justified in Appendix B.

Now add $\omega_m(k)a_\lambda(k)\tilde{\Phi}_m$ on both sides. Since $E^V(m, N)$ is the ground state energy and $\omega_m(k) > 0$ the operator $\tilde{H}_m - E^V(m, N) + \omega_m(k)$ has a bounded inverse $R(\omega_m(k))$ and hence

$$\begin{aligned} a_\lambda(k)\tilde{\Phi}_m &= 2\sqrt{\alpha}R(\omega_m(k)) \frac{\chi_\Lambda(k)}{|k|^{1/2}}\varepsilon_\lambda(k) \cdot \sum_{j=1}^N (p_j + \sqrt{\alpha}\tilde{A}(x_j))(1 - e^{-ik \cdot x_j})\tilde{\Phi}_m \\ &\quad + iR(\omega_m(k))\chi_\Lambda(k) \frac{g}{2}\sqrt{\alpha} \frac{k \wedge \varepsilon_\lambda(k)}{\sqrt{|k|}} \cdot \sum_{j=1}^N \sigma_j e^{-ik \cdot x_j} \tilde{\Phi}_m \\ &\quad - R(\omega_m(k))\omega_m(k)iw_\lambda(k, X)\tilde{\Phi}_m . \end{aligned} \quad (54)$$

For consistency, note that for $|k| > \Lambda$, $w_\lambda(k, X) = 0$. Hence $a_\lambda(k)\tilde{\Phi}_m = 0$, i.e., for these modes $\tilde{\Phi}_m$ is the vacuum as it should be for a minimizer. Since $\|R(\omega_m(k))\| \leq \omega_m(k)^{-1}$ the norm of the last term is bounded by

$$\sqrt{\alpha}\sqrt{N} \frac{\chi_\Lambda(k)}{|k|^{1/2}} \| |X| \tilde{\Phi}_m \| . \quad (55)$$

To bound the norm of the first term we need to estimate

$$\begin{aligned}
 & \left\| \sum_{j=1}^N R(\omega_m(k)) \varepsilon_\lambda(k) \cdot (p_j + \sqrt{\alpha} \tilde{A}(x_j)) (1 - e^{-ik \cdot x_j}) \tilde{\Phi}_m \right\| \\
 &= \sup_{\|\eta\| \leq 1} \left| \sum_{j=1}^N (\varepsilon_\lambda(k) \cdot (p_j + \sqrt{\alpha} \tilde{A}(x_j))) R(\omega_m(k)) \eta, (1 - e^{-ik \cdot x_j}) \tilde{\Phi}_m \right| \\
 &\leq \sup_{\|\eta\| \leq 1} \left[\sum_{j=1}^N \|(p_j + \sqrt{\alpha} \tilde{A}(x_j)) R(\omega_m(k)) \eta\|^2 \right]^{1/2} \left[\sum_{j=1}^N \|(1 - e^{-ik \cdot x_j}) \tilde{\Phi}_m\|^2 \right]^{1/2}
 \end{aligned} \tag{56}$$

Next estimate the square of the first factor to get

$$\begin{aligned}
 & \left(\eta, R(\omega_m(k)) \left[\sum_{j=1}^N (p_j + \sqrt{\alpha} \tilde{A}(x_j))^2 \right] R(\omega_m(k)) \eta \right) \\
 &\leq a \left(\eta, R(\omega_m(k)) H_m^V R(\omega_m(k)) \eta \right) + b \\
 &\leq a \left(\eta, R(\omega_m(k)) \eta \right) + (a E^V(m, N) + b) \left(\eta, R(\omega_m(k))^2 \eta \right) \\
 &\leq C \frac{(\Lambda + 1)}{|k|^2} \quad \text{for } |k| \leq \Lambda
 \end{aligned} \tag{57}$$

where a and b are independent of m . Since $\sup_{m < 1} E^V(m, N) < \infty$ the constant C is also independent of m . Finally the second factor in (56) is bounded by $|k| \|X\| \|\tilde{\Phi}_m\|$. The term containing the Pauli matrices in (54) is estimated similarly. In conjunction with (54), (52), (56) and (57) this shows that

$$\|a_\lambda(k) \Phi_m\| \leq C \sqrt{\alpha} (\Lambda + 1)^{1/2} \frac{\chi_\Lambda(k)}{|k|^{1/2}} \|X\| \|\tilde{\Phi}_m\| \tag{58}$$

This, together with Lemma 6.2, proves the theorem. $\square$

Next we differentiate (54) with respect to k . There is a slight problem with this calculation since the polarization vectors cannot be defined in a smooth fashion globally. We make the following choice for the polarization vectors.

$$\varepsilon_1(k) = \frac{(k_2, -k_1, 0)}{\sqrt{k_1^2 + k_2^2}} \tag{59}$$

and

$$\varepsilon_2(k) = \frac{k}{|k|} \wedge \varepsilon_1(k). \tag{60}$$

6.3. Theorem (Photon derivative bound). *Assume that there is binding, i.e., $\Sigma^V(m, N) - E^V(m, N) > 0$. Assume that Φ_m is a normalized ground state for the many-body Hamiltonian H_m^V , $m \geq 0$. Then for $|k| < \Lambda$ and $(k_1, k_2) \neq (0, 0)$*

$$\|\nabla_k a_\lambda(k) \Phi_m\| < \frac{Q\sqrt{\alpha}(1 + |g|)}{|k|^{1/2}\sqrt{k_1^2 + k_2^2}}, \quad (61)$$

where Q is a finite constant independent of Φ_m , g , α , Λ , and depends on m only through the binding energy $\Sigma^V(m, N) - E^V(m, N) > 0$.

Proof. We differentiate (54) with respect to k and obtain

$$\begin{aligned} \nabla_k(a_\lambda(k))\tilde{\Phi}_m = & 2\sqrt{\alpha}R(\omega_m(k))^2 \frac{k}{|k|} \varepsilon_\lambda(k) \cdot \sum_{j=1}^N (p_j + \sqrt{\alpha} \tilde{A}(x_j)) \frac{(1 - e^{-ik \cdot x_j})}{|k|^{1/2}} \tilde{\Phi}_m + \\ & 2\sqrt{\alpha}R(\omega_m(k)) \nabla_k(\varepsilon_\lambda(k)) \cdot \sum_{j=1}^N (p_j + \sqrt{\alpha} \tilde{A}(x_j)) \frac{(1 - e^{-ik \cdot x_j})}{|k|^{1/2}} \tilde{\Phi}_m + \\ & 2\sqrt{\alpha}R(\omega_m(k)) \varepsilon_\lambda(k) \cdot \sum_{j=1}^N (p_j + \sqrt{\alpha} \tilde{A}(x_j)) \nabla_k \left(\frac{(1 - e^{-ik \cdot x_j})}{|k|^{1/2}} \right) \tilde{\Phi}_m \\ & + \frac{g}{2} \sqrt{\alpha} \nabla_k \left(iR(\omega_m(k)) \frac{k \wedge \varepsilon_\lambda(k)}{\sqrt{|k|}} \cdot \sum_{j=1}^N \sigma_j e^{-ik \cdot x_j} \tilde{\Phi}_m \right) \\ & - \nabla_k (R(\omega_m(k)) \omega_m(k) i w_\lambda(k, X) \tilde{\Phi}_m) . \end{aligned}$$

The norms of the first and third terms can be estimated precisely the same way as in (56) and (57), and yields a bound of the form

$$\frac{C\sqrt{\alpha}}{|k|^{3/2}} \|(1 + |X|) \tilde{\Phi}_m\| . \quad (62)$$

For the second term, a straightforward calculation shows that

$$|\nabla_k \varepsilon_i(k)| \leq \frac{\text{const.}}{\sqrt{k_1^2 + k_2^2}} \quad \text{for } i = 1, 2. \quad (63)$$

The last term is dealt with in a similar fashion as the previous ones. Using the steps in (56) and (57), this leads to the bound

$$\|\nabla_k(a_\lambda(k))\tilde{\Phi}_m\| \leq \frac{C\sqrt{\alpha}}{|k|^{1/2}\sqrt{k_1^2 + k_2^2}} \|(1 + |X|) \tilde{\Phi}_m\| . \quad (64)$$

The fourth term can be estimated in the same fashion to yield a similar result.

Differentiating (51) leads to the same estimate with $\tilde{\Phi}_m$ replaced by Φ_m . This, together with Lemma 6.2, proves the theorem. $\square$

As for the proof of Theorem 6.1, our somewhat formal calculations above are rigorously justified in Appendix B.

7 Proof of Theorem 2.1

The proof will be done in two steps.

Proof. Step 1. The Hamiltonian H_m^V has a normalized ground state Φ_m , by Theorem 4.1. Pick a sequence $m_1 > m_2 > \dots$ tending to zero and denote the corresponding eigenvectors by Φ_j . This sequence is a minimizing sequence for H_0^V by Theorem 5.3. Since $\|\Phi_j\|$ is bounded there is a subsequence (call it again Φ_j) which has a weak limit Φ . Since $H_0^V - E^V(0, N) \geq 0$ and by the lower semi-continuity of non-negative quadratic forms (in our case, $H_0^V - E^V(0, N)$)

$$0 \leq (\Phi, (H_0^V - E^V(0, N))\Phi) \leq \liminf_{j \rightarrow \infty} (\Phi_j, (H_0^V - E^V(0, N))\Phi_j) = 0.$$

Hence Φ will be a (normalized) ground state if we show that $\|\Phi\| = 1$ (i.e. $\Phi_j \rightarrow \Phi$ strongly). It is important to note, however that if we write $\Phi_j = \{\Phi_{j,0}, \Phi_{j,1}, \dots, \Phi_{j,n}, \dots\}$, where $\Phi_{j,n}$ is the n -photon component of Φ_j then it suffices to prove the L^2 norm-convergence of each $\Phi_{j,n}$. The reason is the uniform bound on the total average photon number; see the remark after Theorem 6.1 which implies

$$\sum_{n \geq N} \|\Phi_{j,n}\|^2 \leq \text{const } N^{-1}.$$

Likewise, it suffices to prove the strong L^2 convergence in the bounded domain in which $|X| < R$ for each finite R . The reason for this is the exponential decay given in Lemma 6.2, which is uniform by Lemma 5.4. Finally, by Theorem 6.1, $\Phi_{j,n}(X, k_1, \dots, k_n)$ vanishes if $|k_i| > \Lambda$ for some i . So it suffices to show L^2 convergence for $\Phi_{j,n}$ restricted to

$$\Omega = \{(X, k_1, \dots, k_n) : |X| < R; |k_i| < \Lambda, i = 1, \dots, n\} \subset \mathbb{R}^{3(N+n)}$$

for each $R > 0$.

Step 2. For each $p < 2$ and $R > 0$ we show that $\Phi_{j,n}$ restricted to Ω is a bounded sequence in $W^{1,p}(\Omega)$. The key to this bound is (61) and

$$(a_\lambda(k)\Phi_j)_{n-1}(X, k_1, \dots, k_{n-1}) = \sqrt{n}\Phi_{j,n}(X, k, k_1, \dots, k_{n-1}) \quad (65)$$

where the arguments $\lambda, \lambda_1, \dots, \lambda_{n-1}$ and the spin indices have been suppressed. By the symmetry of $\Phi_{j,n}$, (65), Hölder's inequality and (61)

$$\begin{aligned}
 & \int_{B_R} dX \int_{|k_1|, \dots, |k_n| < \Lambda} dk_1 \dots dk_n \sum_{i=1}^n |\nabla_{k_i} \Phi_{j,n}(X, k_1, \dots, k_n)|^p \\
 &= n^{1-p/2} \int_{B_R} dX \int_{|k_1|, \dots, |k_n| < \Lambda} dk_1 \dots dk_n |\nabla_{k_1} (a_\lambda(k_1) \Phi_j)_{n-1}(X, k_2, \dots, k_n)|^p \\
 &\leq C \int_{|k_1| < \Lambda} dk_1 \left(\int_{B_R} dX \int_{|k_2|, \dots, |k_n| < \Lambda} dk_2 \dots dk_n \right. \\
 &\quad \left. \times |\nabla_{k_1} (a_\lambda(k_1) \Phi_j)_{n-1}(X, k_2, \dots, k_n)|^2 \right)^{p/2} \\
 &\leq C \int_{|k_1| < \Lambda} dk_1 \|\nabla_{k_1} a_\lambda(k_1) \Phi_j\|^p \leq \text{const} \tag{66}
 \end{aligned}$$

independent of j . The constant C depends on all the parameters, but is finite because $|k_i| \leq \Lambda$ in the integration. Similarly, by Hölder's inequality

$$\begin{aligned}
 \|\chi(|X| < R) \nabla_X \Phi_{j,n}\|_p^p &\leq C \|\chi(|X| < R) \nabla_X \Phi_{j,n}\|_2^p \\
 &\leq C (\Phi_j, \sum_{i=1}^N p_i^2 \Phi_j)^{p/2}
 \end{aligned}$$

which is uniformly bounded by Lemma (A.5).

Since the classical derivative of $a_\lambda(k) \Phi_m$ is not defined in all of Ω one has to check that the weak derivative coincides with the classical derivative a.e.. Because of our definitions (59) and (60), the classical derivative is not defined along the 3-axis.

One has to show that

$$\int_{\Omega} \partial_i \psi \Phi_{j,n} = \lim_{\varepsilon \rightarrow 0} \int_{\Omega_\varepsilon} \partial_i \psi \Phi_{j,n} = - \lim_{\varepsilon \rightarrow 0} \int_{\Omega_\varepsilon} \psi \partial_i \Phi_{j,n} \quad i = 1, \dots, 3(N+n).$$

for any test function $\psi \in C_c^\infty(\Omega)$. Here Ω_ε is Ω with an ε cylinder around the 3-axis removed in each k -ball. The first equality is trivial; it is the second equality that has to be checked. This amounts to showing that the boundary term, coming from the integration by parts vanishes in the limit as ε tends to zero. But this follows immediately from Theorem 6.1.

This shows that $\Phi_{j,n}$ as a function of all its $3(N+n)$ variables, is in the Sobolev space $W^{1,p}(\Omega)$ and that $\sup_j \|\Phi_{j,n}\|_{W^{1,p}(\Omega)} < \infty$. Since $\Phi_{n,j}$ converges weakly in $L^2(\Omega)$ it converges weakly in $L^p(\Omega)$ and since the sequence is bounded in $W^{1,p}(\Omega)$, $\nabla \Phi_{n,j}$ converges weakly to $\nabla \Phi_n$.

The Rellich-Kondrachov theorem (see [22] Theorem 8.9) states that such a sequence converges *strongly* in $L^q(\Omega)$ if $1 \leq q < [3p(N+n)/3(N+n) - p]$. The boundedness of Ω is crucial here. For our purposes we need $q = 2$, and hence we have to pick p such that

$$2 > p > \frac{2 \cdot 3(N+n)}{2 + 3(N+n)} \tag{67}$$

579

which is possible for each N and n . We conclude that $\Phi_{j,n} \rightarrow \Phi_n$ strongly in $L^2(\Omega)$ as $j \rightarrow \infty$, for each n and R . This proves the theorem. $\square$

Remark: Theorem 6.3 essentially says that the derivative is almost, but not quite in L^2 . For high dimensions, $3(N+n)$, the required p is as close as we please to 2 if we require $q = 2$, but $p = 2$ is not allowed. The way out of the difficulty was to prove a uniform bound on the number operator and use this to say that it suffices to prove strong convergence for each n separately. With n then fixed, it is possible to find a $p < 2$ that yields $q = 2$. It is, therefore, crucial to have the derivative in every L^p space with $p < 2$. The resolution of the problem of the infrared singularity is thus seen to be a delicate matter.

A Appendix: Localization estimates

In this appendix we collect a few facts which we use several times in this paper. Generally we worry about localizations of Hamiltonians in configuration space. While this is standard for Schrödinger operators it is somewhat more complicated in the presence of the radiation field. This is chiefly due to the problem of localization of photons.

We begin by stating a few well known facts about partitions of unity. Let β denote one of the 2^N subsets of the set of integers $1, 2, \dots, N$. Its complement is denoted by β^c . As shown in [20] there exists a family of smooth functions j_β having the following four properties.

(i)

$$\sum_{\beta} j_{\beta}^2 = 1. \quad (68)$$

(ii) For $\beta \neq \{1, \dots, N\}$ (including the empty set) the j_β 's are homogeneous of degree 0 and live outside the ball of radius R centered at the origin and

$$\text{supp } j_{\beta} \subset \{X : \min_{i \in \beta, j \in \beta^c} (|x_i - x_j|, |x_j|) \geq c|X|\}, \quad (69)$$

where C is some positive constant.

(iii) In the case where $\beta = \{1, \dots, N\}$, j_β is compactly supported.

Corresponding to these electron localizations we define photon localizations.

For given $\beta \neq \{1, \dots, N\}$ consider the function

$$g_1(y; \beta, X) = \prod_{j \in \beta^c} \left(1 - \chi\left(\frac{y - x_j}{P}\right) \right) \quad (70)$$

where χ is a smoothed characteristic function of the unit ball. Define $g_2(y; \beta, X) = 1 - g_1(y, \beta, X)$. In the variable y , the function g_1 is supported away from the particles in β^c while g_2 lives close to the particles in β^c . Next define, for $i = 1, 2$,

$$j_i(y; \beta, X) = \frac{g_i(y; \beta, X)}{\sqrt{g_1(y; \beta, X)^2 + g_2(y; \beta, X)^2}}. \quad (71)$$

Certainly $j_1^2 + j_2^2 = 1$ and a simple computation shows that

$$|\nabla j_i| \leq \frac{\text{const.}}{P}. \quad (72)$$

In the case where $\beta = \{1, \dots, N\}$ the construction of j_1 and j_2 is similar to the above one except that the function g_1 depends on y , is equal to one in a neighborhood of the origin and is compactly supported.

With the help of j_1 and j_2 the photons can now be localized as was done in Sect. 4. Let $U_\beta(X): \mathcal{F} \rightarrow \mathcal{F} \otimes \mathcal{F}$ be the corresponding isometric transformation, i.e., the one that is defined via the relation

$$U_\beta(X)a^\#(h)U_\beta^*(X) = a^\#(j_1h) \otimes \mathcal{I} + \mathcal{I} \otimes a^\#(j_2h). \quad (73)$$

The tensor product indicated is a tensor product between Fock spaces.

We denote by H_β the Hamiltonian of the form (4) with photon mass, but only for the particles in the set β . More precisely this operator acts on $L^2(\mathbb{R}^{3|\beta|}) \otimes \mathcal{F}$. By H^{β^c} we denote the Hamiltonian of the form (4) with photon mass, but only for the particles in the set β^c where the interaction with the nuclei has been dropped. This operator acts on $L^2(\mathbb{R}^{3|\beta^c|}) \otimes \mathcal{F}$. In particular we keep the interaction among those particles. In the case where $\beta = \{1, \dots, N\}$ the Hamiltonian $H^{\beta^c} = H_f(m)$.

A.1. Lemma (Localization of the Hamiltonian). *For every β*

$$j_\beta H j_\beta = U_\beta^*(X) j_\beta [H_\beta \otimes \mathcal{I} + \mathcal{I} \otimes H^{\beta^c}] j_\beta U_\beta(X) + o(1). \quad (74)$$

For $\beta \neq \{1, \dots, N\}$, $o(1) \rightarrow 0$ as first $R \rightarrow \infty$ and then $P \rightarrow \infty$. If $\beta = \{1, \dots, N\}$ then $o(1) \rightarrow 0$ as $P \rightarrow \infty$ for every fixed $R > 0$.

Proof. Our immediate aim is to compare the field energy H_f with the localized field energy $U_\beta^*(X)[H_f \otimes \mathcal{I} + \mathcal{I} \otimes H_f]U_\beta(X)$. For simplicity the various indices are suppressed and $U_\beta(X)$ is replaced by U_β . The variable X plays no role here. Pick an orthonormal basis $\{g_j\}_{j=1}^\infty$ of $L^2(\mathbb{R}^3)$ in $H^{1/2}(\mathbb{R}^3)$. States of the form

$$\zeta = \text{const.} a_{\lambda_{i_1}}^*(g_{i_1}) \cdots a_{\lambda_{i_k}}^*(g_{i_k}) |0\rangle \quad (75)$$

where k is finite, form an orthonormal basis in the Fock space. The field energy acts on such states as

$$H_f \zeta = \sum_{j=1}^k a_{\lambda_{i_1}}^*(g_{i_1}) \cdots a_{\lambda_{i_j}}^*(\omega g_{i_j}) \cdots a_{\lambda_{i_k}}^*(g_{i_k}) |0\rangle. \quad (76)$$

Thus, we have that

$$H_f \zeta = U_\beta^* \sum_{j=1}^k c_{\lambda_{i_1}}^* (g_{i_1}) \cdots c_{\lambda_{i_j}}^* (\omega g_{i_j}) \cdots c_{\lambda_{i_k}}^* (g_{i_k}) U_\beta |0\rangle \quad (77)$$

and

$$H_f = U_\beta^* [H_f \otimes \mathcal{I} + \mathcal{I} \otimes H_f] U_\beta + E_f \quad (78)$$

where the error E_f is given by

$$\begin{aligned} E_f \zeta = U_\beta^* \sum_{j=1}^k c_{\lambda_{i_1}}^* (g_{i_1}) \cdots (a_{\lambda_{i_j}}^* ([j_1, \omega] g_{i_j}) \otimes \mathcal{I} + \mathcal{I} \otimes a_{\lambda_{i_j}}^* ([j_2, \omega] g_{i_j})) \\ \cdots c_{\lambda_{i_k}}^* (g_{i_k}) U_\beta |0\rangle . \end{aligned} \quad (79)$$

Thus E_f is given by the operator (the λ 's are omitted)

$$\begin{aligned} E_f = U_\beta^* \sum_k [a^*([j_1, \omega] g_k) \otimes \mathcal{I} + \mathcal{I} \otimes a^*([j_2, \omega] g_k)] \\ \times [a(j_1 g_k) \otimes \mathcal{I} + \mathcal{I} \otimes a(j_2 g_k)] U_\beta . \end{aligned} \quad (80)$$

The expression for the operator E_f does not look hermitian but it is, remembering that U_β^* is a partial isometry. Standard estimates lead to

$$|(\Psi, E_f \Psi)| \leq (\|[\omega, j_1]\| + \|[\omega, j_2]\|) (\Psi, [\mathcal{N} + 1] \Psi) . \quad (81)$$

where $\mathcal{N}$ is the number operator.

Here $\| [j_1, \omega] \|$ denotes the operator norm associated with the kernel $[j_1, \omega]$. This norm can be estimated using the formula

$$[j_1, \omega] = [j_1, \omega^2] \frac{1}{\omega} + \omega^2 \left[j_1, \frac{1}{\omega} \right] . \quad (82)$$

Recalling the definition of j_1 , the operator norm of the first term is easily seen to be bounded by a *const.*/ P . Likewise, the second term, using the formula

$$\frac{1}{\sqrt{p^2 + m^2}} = \frac{1}{\pi} \int_0^\infty \frac{1}{t + p^2 + m^2} \frac{dt}{\sqrt{t}} , \quad (83)$$

can be estimated by *const.*/ P . The term $\| [j_2, \omega] \|$ is estimated in a similar fashion. The estimate (81) immediately shows that for a general state Φ we have that

$$|(\Phi, [H_f - U_\beta^* [H_f \otimes \mathcal{I} + \mathcal{I} \otimes H_f] U_\beta] \Phi)| \leq \frac{\text{const.}}{P} (\Phi, \mathcal{N} \Phi) . \quad (84)$$

Since the photons have a mass we can estimate the number operator in terms of the field energy. The field energy is relatively bounded with respect to

the Hamiltonian, i.e., $H_f \leq aH_m^V + b$ for some positive constants a and b , and thus we obtain

$$|(\Phi, [H_f - U_\beta^*[H_f \otimes \mathcal{I} + \mathcal{I} \otimes H_f]U_\beta]\Phi)| \leq \frac{\text{const.}}{Pm} (\Phi, [aH_m^V + b]\Phi). \quad (85)$$

Note that this estimate had nothing to do with the electron, in particular the x -space cutoff is not present in the calculation.

Next we have to compare $\sum_{j=1}^N (p_j + \sqrt{\alpha} A(x_j))^2$ with

$$U_\beta^*(X) j_\beta \left[\sum_{i \in \beta} (p_i + \sqrt{\alpha} A(x_i))^2 \otimes \mathcal{I} + \mathcal{I} \otimes \sum_{j \in \beta^c} (p_j + \sqrt{\alpha} A(x_j))^2 \right] j_\beta U_\beta(X).$$

This time the X -space cutoff is important. We would like to estimate the difference

$$\begin{aligned} & j_\beta \sum_{i \in \beta} [(p_i + \sqrt{\alpha} A(x_i))^2 - U_\beta^*(X)(p_i + \sqrt{\alpha} A(x_i))^2 \otimes \mathcal{I} U_\beta(X)] j_\beta \\ & + j_\beta \sum_{i \in \beta^c} [(p_i + \sqrt{\alpha} A(x_i))^2 - U_\beta^*(X) \mathcal{I} \otimes (p_i + \sqrt{\alpha} A(x_i))^2 U_\beta(X)] j_\beta. \end{aligned} \quad (86)$$

It suffices to treat the first term, the other is similar. It can be easily expressed as

$$j_\beta \left[\sum_{i \in \beta} (p_i + \sqrt{\alpha} A(x_i)) Q_i + Q_i (p_i + \sqrt{\alpha} A(x_i)) - Q_i^2 \right] j_\beta \quad (87)$$

where

$$Q_i = p_i + \sqrt{\alpha} A(x_i) - U_\beta^*(X)(p_i + \sqrt{\alpha} A(x_i)) \otimes \mathcal{I} U_\beta(X). \quad (88)$$

Using the form boundedness of the kinetic energy with respect to the full Hamiltonian, we have

$$\left(\Psi, \sum_{j=1}^N (p_j + \sqrt{\alpha} A(x_j))^2 \Psi \right) \leq a(\Psi, H_m^V \Psi) + b(\Psi, \Psi) \quad (89)$$

for positive constants a and b . Thus, using Schwarz' inequality it suffices to show that

$$\|Q_i j_\beta \Psi\| = o(1) \text{ for } i \in \beta, \quad (90)$$

as R (the localization radius for the electrons) tends to infinity. Denote by

$$h_{i,x}^\lambda(y) = (2\pi)^{-3/2} \int_{|k| < \Lambda} \frac{1}{\omega(k)} \varepsilon_i^\lambda(k) e^{ik \cdot (y-x)} d^3 k. \quad (91)$$

Explicitly, Q_i is given by

$$\begin{aligned} & p_i - U_\beta^*(X) p_i \otimes \mathcal{I} U_\beta(X) \\ & + U_\beta^*(X) \left[\sum_\lambda a_\lambda([j_1 - 1]h_x^\lambda) \otimes \mathcal{I} + \mathcal{I} \otimes a_\lambda(j_2 h_x^\lambda) \right] U_\beta(X) \\ & + U_\beta^*(X) \left[\sum_\lambda a_\lambda^*([j_1 - 1]h_x^\lambda) \otimes \mathcal{I} + \mathcal{I} \otimes a_\lambda^*(j_2 h_x^\lambda) \right] U_\beta(X), \quad (92) \end{aligned}$$

and it suffices to estimate each of these terms separately. Each of the last two terms can be brought into the form

$$\|U_\beta^*(X) a^\#(f) \otimes \mathcal{I} U_\beta(X) j_\beta \Psi\| \quad (93)$$

where f is one of the functions

$$[j_1(y, \beta, X) - 1]h_{1,x_j}^\lambda(y) \text{ or } j_2(y, \beta, X)h_{1,x_j}^\lambda(y) \quad j \in \beta. \quad (94)$$

The terms (93) are estimated by

$$\sup_X \{j_\beta(X) \| [j_1 - 1]h_{i,x_j}^\lambda \|_2\} \sqrt{(\Psi, (\mathcal{N} + 1)\Psi)} \quad (95)$$

respectively

$$\sup_X \{j_\beta(X) \| j_2 h_{1,x_j}^\lambda \|_2\} \sqrt{(\Psi, (\mathcal{N} + 1)\Psi)}. \quad (96)$$

In both formulas the index j is in β . The function j_β lives in the region where $|x_i - x_j| \geq cR$ for $i \in \beta$ and $j \in \beta^c$. The function $j_1 - 1$ (and likewise j_2) is not zero only if $|y - x_j| \leq P$ for some $j \in \beta^c$. Thus, $j_\beta(X)(j_1 - 1)(y)$ and $j_\beta(X)j_2(y)$ are nonzero only if $|y - x_i| \geq cR - P$. As $cR - P$ gets large only the tail of the function h^λ contributes to the integral which can be made as small as we please. The number operator is bounded by the field energy times $1/m$ which in turn is bounded by the full energy.

To estimate the first term in (92) we calculate

$$\begin{aligned} & p_i - U_\beta^*(X) p_i \otimes \mathcal{I} U_\beta(X) = \\ & U_\beta^*(X) \sum_k [a^*([p_i, j_1]g_k) \otimes \mathcal{I} + \mathcal{I} \otimes a^*([p_i, j_2]g_k)] \times \\ & [a(j_1 g_k) \otimes \mathcal{I} + \mathcal{I} \otimes a(j_2 g_k)] U_\beta(X). \quad (97) \end{aligned}$$

Note that the tensor product in the first line is different from the second. In the first the identity acts on $L^2(\mathbb{R}^{3|\beta^c|}) \otimes \mathcal{F} \otimes \mathcal{F}$ while in the second $\otimes$ indicates the tensor product of the Fock spaces only. The functions g_k

indicates a basis of $L^2(\mathbb{R}^3)$. The operators $U_\beta^*(X)$ and $U_\beta(X)$ have unit norm. Thus

$$\begin{aligned} & \| U_\beta^*(X) \sum_k [a^*([p_i, j_1]g_k) \otimes \mathcal{I} + \mathcal{I} \otimes a^*([p_i, j_2]g_k)] \times \\ & [a(j_1g_k) \otimes \mathcal{I} + \mathcal{I} \otimes a(j_2g_k)] U_\beta(X) \Psi \| \\ & \leq (\| [p_i, j_1] \| + \| [p_i, j_2] \|) \| \sqrt{\mathcal{N} + 1} \Psi \|, \end{aligned} \quad (98)$$

where $\| \cdot \|$ indicates that the operator norm has been taken. The norms of the commutators are of the order $1/P$ and hence vanish as $P \rightarrow \infty$. Since the photons have a mass we can estimate the number operator in terms of the field energy.

Similar consideration apply to the β^c term in (86). The only difference is that instead of (95) and (96) we have

$$\sup_X \{ j_\beta(X) \| j_1 h_{x_j}^\lambda \|_2 \} \sqrt{(\Psi, (\mathcal{N} + 1) \Psi)} \quad (99)$$

respectively

$$\sup_X \{ j_\beta(X) \| [j_2 - 1] h_{x_j}^\lambda \|_2 \} \sqrt{(\Psi, (\mathcal{N} + 1) \Psi)}, \quad (100)$$

with $j \in \beta^c$. Again this terms tend to zero as $P \rightarrow \infty$. The proof for the case where $\beta = \{1, \dots, N\}$ is similar but simpler since the operator U_β does not depend on X .

Finally, we have to compare the $\sigma \cdot B$ term with its localized counterparts. The estimates are similar to, but much easier than the estimates for $(p + \sqrt{\alpha} A(x))^2$ and are omitted for the convenience of the reader and authors who, by now, are exhausted. $\square$

A simple consequence of Lemma A.1 is the following.

A.2. Corollary. *Let ϕ be a smooth function on $\mathbb{R}^{3N}$ such that $j_\beta \phi \equiv 0$ for $\beta = \{1, \dots, N\}$. Thus, ϕ depends on R . Then, as operators,*

$$\phi H \phi \geq (\Sigma^V(m, N) + o(1)) \phi^2. \quad (101)$$

Here, $\Sigma^V(m, N) = \min_{1 \leq N' < N} (E^V(N') + E^0(N - N'))$ and $o(1)$ vanishes as $R \rightarrow \infty$.

Proof. By the IMS localization formula we have that

$$\phi H \phi = \sum_\beta \phi j_\beta H j_\beta \phi - \phi^2 \sum_\beta |\nabla j_\beta|^2, \quad (102)$$

where the second term goes to 0 as $R \rightarrow \infty$. With our assumption on ϕ only the sets β with $\beta^c \neq \emptyset$ contribute. From Lemma A.1 we get that

$$\phi H \phi = \sum_{\beta} U_{\beta}^*(X) \phi j_{\beta} [H_{\beta} \otimes \mathcal{I} + \mathcal{I} \otimes H^{\beta^c}] j_{\beta} \phi U_{\beta}(X) + o(1) \quad (103)$$

as first $R \rightarrow \infty$ then $P \rightarrow \infty$. Certainly $H_{\beta} \geq E^V(m, |\beta|)$ and $H^{\beta^c} \geq E^0(m, |\beta^c|)$ from which the statement immediately follows.

A.3. Lemma. *Let Ψ_n be a normalized sequence in $\mathcal{H}$ whose energy is uniformly bounded and such that for any $\Phi \in \mathcal{H}$ with finite energy,*

$$(\Psi_n, \Phi) \rightarrow 0, \text{ and } (\Psi_n, H\Phi) \rightarrow 0. \quad (104)$$

Then

$$(\phi_R \Psi_n, U^* \mathcal{I} \otimes P_2 U \phi_R \Psi_n) \rightarrow 0. \quad (105)$$

Here U is the Fock space localization U_{β} that corresponds to $\beta = \{1, \dots, N\}$.

Proof. Since the energy of Ψ_n is uniformly bounded we also know that

$$(\Psi_n, H^0(m) \Psi_n) \leq C. \quad (106)$$

is uniformly bounded

Let us describe the operator $\mathcal{I} \otimes P_2 U$ in more detail. Recall that

$$U a^*(h_{i_1}) \cdots a^*(h_{i_k}) |0\rangle = c^*(h_{i_1}) \cdots c^*(h_{i_k}) U |0\rangle \quad (107)$$

where $|0\rangle$ denotes the vacuum vector in Fock space and $U|0\rangle = |0\rangle \otimes |0\rangle$. Hence, using the definition of $c^*(h)$, we find that

$$\mathcal{I} \otimes P_2 U a^*(h_{i_1}) \cdots a^*(h_{i_k}) |0\rangle = a^*(h_{i_1} j_1) \cdots a^*(h_{i_k} j_1) |0\rangle \otimes |0\rangle. \quad (108)$$

The projection P_2 annihilates the photons in the second factor. In other words, the operator $\mathcal{I} \otimes P_2 U$ when acting on a state

$$\Psi = \{\Psi^0, \Psi^1(y_1), \Psi^2(y_1, y_2), \dots\} \quad (109)$$

produces the localized state $\Gamma(j_1) \Psi \otimes |0\rangle$ where

$$\Gamma(j_1) \Psi = \{\Psi^0, j_1(y_1) \Psi^1(y_1), j_1(y_1) j_1(y_2) \Psi^2(y_1, y_2), \dots\}. \quad (110)$$

It follows that

$$\begin{aligned} (\phi_R \Psi_n, U^* \mathcal{I} \otimes P_2 U \phi_R \Psi_n) &= \|\mathcal{I} \otimes P_2 U \phi_R \Psi_n\|^2 \\ &= \|\Gamma(j_1) \phi_R \Psi_n\|^2 \end{aligned}$$

Next, we show that (106) implies that

$$\Gamma(j_1) \phi_R \Psi_n \rightarrow 0. \quad (111)$$

To achieve that we note first that on account of the positive mass we have that $(\Psi_n, \mathcal{N}\Psi_n)$ is uniformly bounded. Since Ψ_n is of the form

$$\{\Psi_n^0, \Psi_n^1(X, y_1), \Psi_n^2(X, y_1, y_2), \dots\}$$

we know that $\sum_{k \geq M} (\Psi_n^k, \Psi_n^k) \leq \text{const}/M$. It is therefore sufficient to prove (111) for each function

$$\Psi_n^M(X, y_1, \dots, y_k) .$$

From the lemma below we learn that

$$\sum_{j=1} (\Psi_n, p_j^2 \Psi_n) \tag{112}$$

is uniformly bounded. Thus, we can write (111) as

$$\Gamma(j_1) \phi_R \left(1 + \sum_{j=1} p_j^2 + H_f\right)^{-1/2} \left(1 + \sum_{j=1} p_j^2 + H_f\right)^{1/2} \Psi_n \tag{113}$$

which vanishes as $n \rightarrow \infty$ since $\|(1 + \sum_{j=1} p_j^2 + H_f)^{1/2} \Psi_n\|$ is uniformly bounded and since

$$\Gamma(j_1) \phi_R \left(1 + \sum_{j=1} p_j^2 + H_f\right)^{-1/2}$$

is compact on every finite particle subspace. Compactness follows from the fact that for continuous functions f and g vanishing at infinity the operator $f(i\nabla)g(x)$ is compact.

A.4. Lemma (Bound on $A(x)^2$). *For each $x \in \mathbb{R}^3$ and ultraviolet cutoff Λ write $A(x) = D(x) + D^*(x)$ where D contains the annihilation operators in $A(x)$ and D^* the creation operators. Similarly, write $B(x) = E(x) + E^*(x)$. As operator bounds*

$$\begin{aligned} H_f &\geq \frac{1}{8\pi\Lambda} D^*(x) D(x) \\ H_f + \frac{\Lambda}{2} &\geq \frac{1}{8\pi\Lambda} D(x) D^*(x) \\ H_f + \frac{\Lambda}{8} &\geq \frac{1}{32\pi\Lambda} A(x)^2 \\ H_f &\geq \frac{3}{8\pi\Lambda^3} E^*(x) E(x) \\ H_f + \frac{3\Lambda}{4} &\geq \frac{3}{8\pi\Lambda^3} E(x) E^*(x) \\ H_f + \frac{3\Lambda}{16} &\geq \frac{3}{32\pi\Lambda^3} B(x)^2 . \end{aligned} \tag{114}$$

Proof. We write $A(x) = D(x) + D^*(x)$ with

$$D(x) = \sum_{\lambda} \int_{|k| < \Lambda} |k|^{-1/2} \varepsilon_{\lambda}(k) \exp[ik \cdot x] a_{\lambda}(k) d^3 k.$$

There are thus four terms in $A(x)^2$. Using the Schwarz inequality, the (DD) term can be bounded above by $(D^*D)/2 + (DD^*)/2$. On the other hand, $(DD^*) = (D^*D) + \Gamma$, where Γ is the commutator $\int 2/|k| = 4\pi\Lambda^2$; the factor 2 comes from the two polarizations $\lambda = 1, 2$. Altogether, we obtain

$$A(x)^2 \leq 4D^*(x)D(x) + 4\pi\Lambda^2.$$

Finally, we use the Schwarz inequality again to obtain

$$\begin{aligned} & \sum_{\lambda} \int \overline{h_{\lambda}(k)} a_{\lambda}^*(k) d^3 k \sum_{\lambda} \int h_{\lambda}(k) a_{\lambda}(k) d^3 k \\ & \leq \sum_{\lambda} \int |h_{\lambda}(k)|^2 / |k| d^3 k \sum_{\lambda} \int |k| a_{\lambda}^*(k) a_{\lambda}(k) d^3 k. \end{aligned}$$

In our case, $h_{\lambda}(k) = \varepsilon_{\lambda}(k) \exp[ik \cdot x] / \sqrt{|k|}$, so $\sum_{\lambda} \int |h_{\lambda}(k)|^2 / |k| d^3 k = 8\pi\Lambda$.

For $B = \text{curl} A$, replace Γ by $2\pi\Lambda^4$ and replace $|h_{\lambda}(k)|$ by $\sqrt{|k|}$. $\square$

As a corollary of Lemma A.4 we have the following.

A.5. Lemma (Bound on $(p + A(x))^2$). *For any $\varepsilon > 0$ there are constants $\delta(\varepsilon) > 0$ and $C(\varepsilon) < \infty$ such that*

$$\sum_{j=1}^N \left\{ (p_j + \sqrt{\alpha} A(x_j))^2 + \frac{g}{2} \sqrt{\alpha} \sigma_j \cdot B(x_j) \right\} + \varepsilon H_f \geq \delta(\varepsilon) \sum_{j=1}^N p_j^2 - C(\varepsilon). \quad (115)$$

The constants $\delta(\varepsilon)$, $C(\varepsilon)$ depend on α , g , Λ , N .

Proof. In addition to Lemma A.4, use the facts that for any $0 < \mu, \nu < 1$, $(p_j + \sqrt{\alpha} A(x_j))^2 \geq (1 - \mu)p^2 + (1 - 1/\mu)\alpha A(x_j)^2$ and $2\sigma_j \cdot B(x_j) \geq -\nu B(x_j)^2 - 1/\nu$. $\square$

B Appendix: Verification of infrared bounds

The proofs of the infrared bounds in Sect. 6 are somewhat formal. In particular, we carried out the calculations tacitly assuming that $a_{\lambda}(k)\tilde{\Phi}_m$ (which is itself only defined for almost every k) is in the domain of the Operator $\tilde{H}_m$. One can actually prove this when $\tilde{H}_m$ is self-adjointly realized in terms of the Friedrichs' extension and thereby make all the formal computations in Sect. 6 rigorous. Instead of doing so, we give here alternative proofs of the

Theorems in Sect. 6 which avoid any reference to a domain of $\tilde{H}_m$. All the arguments can be carried out on the level of quadratic forms.

We recall that $\tilde{H}_m$ is the Hamiltonian H_m^V after an “operator-valued gauge transformation”. Our remarks here about quadratic forms in relation to $\tilde{H}_m$ could just as well be applied to H_m^V itself.

In order to keep the notation simple, we give the proof of the infrared bounds for the case of a single charged particle ($N = 1$) with no magnetic moment, i.e., $g = 0$. There is no difficulty in deriving these bounds for the general case.

Denote by $\mathcal{S}$ the set of all finite linear combinations of vectors that are products of $C_c^\infty(\mathbb{R}^3)$ -functions and states in $\mathcal{F}$ that have only a finite number of photons. It is well known that this set is dense in $\mathcal{H}$, and that the quadratic form $(\Psi, \Psi)_+ := (\Psi, (\tilde{H}_m - E^V(m, 1) + 1)\Psi)$ is defined for all Ψ in $\mathcal{S}$ and is bounded below by $\|\Psi\|^2$. Hence this quadratic form is closable and the closure of $\mathcal{S}$ in this inner product is a Hilbert space $Q(\tilde{H}_m)$ with inner product $(\cdot, \cdot)_+$ and norm $\|\Psi\|_+ = \sqrt{(\Psi, \Psi)_+}$.

An eigenfunction $\tilde{\Phi}_m$ of $\tilde{H}_m$ in the weak sense is a vector in $Q(\tilde{H}_m)$ such that

$$(\Psi, \tilde{\Phi}_m)_+ = e(\Psi, \tilde{\Phi}_m) \tag{116}$$

for some real number e and for all $\Psi \in Q(\tilde{H}_m)$. It is in this sense that we proved in Sect. 4 that a ground state exists for the model with massive photons. (This implies that $\tilde{\Phi}_m$ is in an eigenstate of the Friedrichs’ extension of $\tilde{H}_m$).

Define the smeared operators

$$a(f) = \sum_{\lambda} \int a_{\lambda}(k) \overline{f(k, \lambda)} dk, \tag{117}$$

where $f(k, \lambda)$ is any function in $L^2(\mathbb{R}^3; \mathbb{C}^2)$. It is not difficult to show that $a(f)\tilde{\Phi}_m$ is in the form domain of $\tilde{H}_m$. To this end define $a_R(f) = R[\mathcal{N} + R]^{-1}a(f)$. Here R is some large real number (which we eventually take towards infinity) and $\mathcal{N}$ is the number operator.

It is straightforward to see that $a_R(f)$ and $a_R^*(f)$ are bounded operators on $Q(\tilde{H}_m)$ for every $R > 0$, i.e.,

$$\|a_R(f)\Psi\|_+ \leq C(R)\|\Psi\|_+, \tag{118}$$

and similarly for $a_R^*(f)$.

Generally, the constant $C(R)$ tends to ∞ as R tends to ∞ . For an eigenfunction of $\tilde{H}_m$, however, this is not the case. Simple but tedious commutator estimates reveal that for any eigenfunction $\tilde{\Phi}_m$ there exists a constant C independent of R such that

$$(a_R(f)\tilde{\Phi}_m, a_R(f)\tilde{\Phi}_m)_+ \leq C(a_R(f)\tilde{\Phi}_m, a_R(f)\tilde{\Phi}_m). \tag{119}$$

The point is that $(a_R(f)\tilde{\Phi}_m, a_R(f)\tilde{\Phi}_m)_+ = (a_R^*(f)a_R(f)\tilde{\Phi}_m, \tilde{\Phi}_m)_+$ plus terms that are uniformly bounded in R . By the previous statement we know that $a_R^*(f)a_R(f)\tilde{\Phi}_m$ is in $Q(\tilde{H}_m)$ and hence

$$\begin{aligned} (a_R^*(f)a_R(f)\tilde{\Phi}_m, \tilde{\Phi}_m)_+ &= e(a_R^*(f)a_R(f)\tilde{\Phi}_m, \tilde{\Phi}_m) \\ &= e(a_R(f)\tilde{\Phi}_m, a_R(f)\tilde{\Phi}_m). \end{aligned} \quad (120)$$

The last expression, however, is bounded uniformly in R , since the condition $\tilde{\Phi}_m \in Q(\tilde{H}_m)$ implies that the expectation value of the field energy in $\tilde{\Phi}_m$ is finite which in turn bounds the last expression in (120). Here we use the fact that the photons have a mass.

From this it follows easily that for a subsequence of R 's tending to infinity, $a_R(f)\tilde{\Phi}_m$ has a weak limit in $Q(\tilde{H}_m)$. Since $a_R(f)\tilde{\Phi}_m \rightarrow a(f)\tilde{\Phi}_m$ strongly this shows that $a(f)\tilde{\Phi}_m \in Q(\tilde{H}_m)$.

Proof of Theorem 6.1. We shall use the abbreviation

$$\sum_{\lambda} \int \dots dk = \oint \dots dk. \quad (121)$$

For our special choice of gauge

$$\tilde{A}^i(x) = a(G^i) + a^*(G^i), \quad i = 1, 2, 3, \quad (122)$$

where we set

$$G_{\lambda}^i(k, x) = \varepsilon_{\lambda}^i(k) |k|^{-1/2} (e^{ik \cdot x} - 1) \chi_{\Lambda}(k). \quad (123)$$

Next, pick any Ψ in $\mathcal{H}$ and calculate (recalling the definition of w in Sect. 6 equation (48))

$$\begin{aligned} (\Psi, (\tilde{H}_m - E^V(m, 1))a(f)\tilde{\Phi}_m) &= -2(\Psi, (f, G)(p + \tilde{A})\tilde{\Phi}_m) \\ &\quad - (\Psi, a(\omega f)\tilde{\Phi}_m) + i(\Psi, (f, \omega w)\tilde{\Phi}_m) \end{aligned} \quad (124)$$

with $\omega(k) = \sqrt{k^2 + m^2}$. This extends, using an approximation argument, to all $\Psi \in Q(\tilde{H}_m)$ and, in particular, to $a(f)\tilde{\Phi}_m$. Here we note that, on account of Lemma A.4 and the assumption on the potential, $\Psi \in Q(\tilde{H}_m)$ implies that $(p + \tilde{A})\Psi \in \mathcal{H}$. Hence

$$\begin{aligned} 0 &\leq (a(f)\tilde{\Phi}_m, (\tilde{H}_m - E^V(m, 1))a(f)\tilde{\Phi}_m) \\ &= -2(a(f)\tilde{\Phi}_m, (f, G)(p + \tilde{A})\tilde{\Phi}_m) \\ &\quad - (a(f)\tilde{\Phi}_m, a(\omega f)\tilde{\Phi}_m) + i(a(f)\tilde{\Phi}_m, (f, \omega w)\tilde{\Phi}_m), \end{aligned} \quad (125)$$

which yields the inequality

$$\begin{aligned} (a(f)\tilde{\Phi}_m, a(\omega f)\tilde{\Phi}_m) &\leq -2(a(f)\tilde{\Phi}_m, (f, G)(p + \tilde{A})\tilde{\Phi}_m) \\ &\quad + i(a(f)\tilde{\Phi}_m, (f, \omega w)\tilde{\Phi}_m) \end{aligned} \quad (126)$$

for all f in $L^2(\mathbb{R}^3; \mathbb{C}^2)$. Pick f of the form $\omega(k)^{-1/2}q(k, \lambda)g_i(k, \lambda)$ where g_i is an orthonormal basis of $L^2(\mathbb{R}^3; \mathbb{C}^2)$ and $q(k, \lambda)$ a bounded function. Summing over this basis, we get on the left side of (126)

$$\sum_i (a(\omega^{-1/2}qg_i)\tilde{\Phi}_m, a(\omega^{1/2}qg_i)\tilde{\Phi}_m) = \int |q(k, \lambda)|^2 \|a_\lambda(k)\tilde{\Phi}_m\|^2 dk, \quad (127)$$

and on the right side

$$-2(a(\omega^{-1}|q|^2G)\tilde{\Phi}_m, (p + \tilde{A})\tilde{\Phi}_m) + i(a(|q|^2w)\tilde{\Phi}_m, \tilde{\Phi}_m). \quad (128)$$

Hence

$$\begin{aligned} \int |q(k, \lambda)|^2 \|a_\lambda(k)\tilde{\Phi}_m\|^2 dk &\leq -2(a(\omega^{-1}|q|^2G)\tilde{\Phi}_m, (p + \tilde{A})\tilde{\Phi}_m) \\ &\quad + i(a(|q|^2w)\tilde{\Phi}_m, \tilde{\Phi}_m). \end{aligned} \quad (129)$$

The right side can be written as

$$\begin{aligned} -2 \int \frac{|q(k, \lambda)|^2}{\omega(k)} (a_\lambda(k)\tilde{\Phi}_m, G_\lambda(k)(p + \tilde{A})\tilde{\Phi}_m) dk \\ + i \int |q(k, \lambda)|^2 (a_\lambda(k)\tilde{\Phi}_m, w_\lambda\tilde{\Phi}_m) dk, \end{aligned} \quad (130)$$

and, applying Schwarz's inequality, this is bounded above by

$$\begin{aligned} 2 \left[\int |q(k, \lambda)|^2 \|a_\lambda(k)\tilde{\Phi}_m\|^2 dk \right]^{1/2} \times \\ \left[\int_{|k| \leq \Lambda} \omega(k)^{-2} |q(k, \lambda)|^2 \|G_\lambda(p + \tilde{A})\tilde{\Phi}_m\|^2 dk \right]^{1/2} \\ + \left[\int_{|k| \leq \Lambda} |q(k, \lambda)|^2 \|w_\lambda\tilde{\Phi}_m\|^2 dk \right]^{1/2} \Big]. \end{aligned} \quad (131)$$

Hence we obtain the bound

$$\begin{aligned} \int |q(k, \lambda)|^2 \|a(k)\tilde{\Phi}_m\|^2 dk &\leq \\ 8 \left[\int_{|k| \leq \Lambda} \omega(k)^{-2} |q(k, \lambda)|^2 \|G_\lambda(k) \cdot (p + \tilde{A})\tilde{\Phi}_m\|^2 dk \right. \\ &\quad \left. + \int_{|k| \leq \Lambda} |q(k, \lambda)|^2 \|w_\lambda\tilde{\Phi}_m\|^2 dk \right]. \end{aligned} \quad (132)$$

Since, $\operatorname{div}_x G_\lambda = 0$ we have that $G_\lambda \cdot (p + \tilde{A})\tilde{\Phi}_m = (p + \tilde{A}) \cdot G_\lambda \tilde{\Phi}_m$. Moreover, $(p + \tilde{A})^2$ is relatively form bounded with respect to $\tilde{H}_m$. But, as in the proof of exponential decay (Lemma 6.2), we have for each $i = 1, 2, 3$

$$(G_\lambda^i \tilde{\Phi}_m, (\tilde{H}_m - E^V(m, 1))G_\lambda^i \tilde{\Phi}_m) = (\tilde{\Phi}_m, |\nabla_x G_\lambda^i|^2 \tilde{\Phi}_m) \quad (133)$$

and we arrive at the bound

$$\begin{aligned} & \int |q(k, \lambda)|^2 \|a_\lambda(k) \tilde{\Phi}_m\|^2 dk \leq \\ & C \int_{|k| \leq \Lambda} \frac{|q(k, \lambda)|^2}{\omega(k)^2} \left[\|G_\lambda \tilde{\Phi}_m\|^2 + \|\nabla_x G_\lambda \tilde{\Phi}_m\|^2 + \omega(k)^2 \|w_\lambda \tilde{\Phi}_m\|^2 \right] dk, \end{aligned} \quad (134)$$

where C is some constant independent of m . Since $q(k, \lambda)$ is arbitrary we obtain for almost every k and each λ that

$$\begin{aligned} & \|a_\lambda(k) \tilde{\Phi}_m\|^2 \leq \\ & C \omega(k)^{-2} \left[\|G_\lambda \tilde{\Phi}_m\|^2 + \|\nabla_x G_\lambda \tilde{\Phi}_m\|^2 + \omega(k)^2 \|w_\lambda \tilde{\Phi}_m\|^2 \right] \chi_\Lambda(k). \end{aligned} \quad (135)$$

The right side is bounded by

$$\frac{C}{|k|} \|\tilde{\Phi}_m\|^2 \chi_\Lambda(k), \quad (136)$$

which is finite on account of the exponential decay of $\tilde{\Phi}_m$. $\square$

Proof of Theorem 6.3. First some notation: For any function $f(k)$ define

$$(\Delta_h f)(k) = f(k+h) - f(k), \quad (137)$$

and

$$\Delta_{-h} a(f) = a(\Delta_h f) \quad (138)$$

Returning to (126) with f replaced by $\Delta_h f$ we have

$$\begin{aligned} & (\Delta_{-h} a(f) \tilde{\Phi}_m, a(\omega \Delta_h f) \tilde{\Phi}_m) \leq \\ & -2(\Delta_{-h} a(f) \tilde{\Phi}_m, (\Delta_h f, G)(p + \tilde{A}) \tilde{\Phi}_m) + i(\Delta_{-h} a(f) \tilde{\Phi}_m, (\Delta_h f, \omega w) \tilde{\Phi}_m) \end{aligned} \quad (139)$$

which can be rewritten as

$$\begin{aligned} & (\Delta_{-h} a(f) \tilde{\Phi}_m, \Delta_{-h} a(\omega f) \tilde{\Phi}_m) \leq (\Delta_{-h} a(f) \tilde{\Phi}_m, a((\Delta_h \omega) f(\cdot + h)) \tilde{\Phi}_m) \\ & -2(\Delta_{-h} a(f) \tilde{\Phi}_m, (f, \Delta_{-h} G)(p + \tilde{A}) \tilde{\Phi}_m) \\ & + i(\Delta_{-h} a(f) \tilde{\Phi}_m, (f, \Delta_{-h}(\omega w)) \tilde{\Phi}_m). \end{aligned} \quad (140)$$

Notice that without the first term on the right of the inequality sign, the structure of this inequality is the same as (126), except that, of course, $\Delta_{-h}a(f)$ plays the role of $a(f)$, $\Delta_{-h}G$ plays the role of G and $\Delta_{-h}(\omega w)$ plays the role of ωw . Thus, without this term we would obtain immediately the estimate analogous to (134),

$$\begin{aligned} & \int |q(k, \lambda)|^2 \|(\Delta_{-h}a_\lambda)(k)\tilde{\Phi}_m\|^2 dk \\ & \leq C \int \frac{|q(k, \lambda)|^2}{\omega(k)^2} \left[\|\Delta_{-h}G_\lambda\tilde{\Phi}_m\|^2 \right. \\ & \quad \left. + \|\nabla_x \Delta_{-h}G_\lambda|\tilde{\Phi}_m\|^2 + \omega^2 \|\Delta_{-h}(\omega w_\lambda)\tilde{\Phi}_m\|^2 \right] dk. \end{aligned} \quad (141)$$

The remaining term in equation (140), after summing over the functions $qg_i/\sqrt{\omega}$, turns into

$$\int ((\Delta_{-h}a_\lambda)(k)\tilde{\Phi}_m, a_\lambda(k-h)\tilde{\Phi}_m) \frac{|q(k, \lambda)|^2}{\omega(k)} (\Delta_h\omega)(k-h) dk \quad (142)$$

which, by Schwarz's inequality, is bounded above by

$$\begin{aligned} & \left[\int |q(k, \lambda)|^2 \|(\Delta_{-h}a_\lambda)(k)\tilde{\Phi}_m\|^2 dk \right]^{1/2} \\ & \times \left[\int \|a_\lambda(k-h)\tilde{\Phi}_m\|^2 \frac{|q(k, \lambda)|^2}{\omega(k)^2} |\Delta_h\omega(k-h)|^2 dk \right]^{1/2}. \end{aligned} \quad (143)$$

This, together with (141), yields

$$\begin{aligned} & \int |q(k, \lambda)|^2 \|(\Delta_{-h}a_\lambda)(k)\tilde{\Phi}_m\|^2 dk \\ & \leq C \int \frac{|q(k, \lambda)|^2}{\omega(k)^2} \left[\|\Delta_{-h}G_\lambda\tilde{\Phi}_m\|^2 + \|\nabla_x \Delta_{-h}G_\lambda|\tilde{\Phi}_m\|^2 \right. \\ & \quad \left. + \omega(k)^2 \|\Delta_{-h}(\omega w_\lambda)\tilde{\Phi}_m\|^2 \right] dk \\ & \quad + C \int \frac{|q(k, \lambda)|^2}{\omega(k)^2} \|a_\lambda(k-h)\tilde{\Phi}_m\|^2 |\Delta_h\omega(k-h)|^2 dk. \end{aligned} \quad (144)$$

Again, since q is arbitrary we obtain for every fixed λ

$$\begin{aligned} \left\| (\Delta_{-h}a_\lambda)(k)\tilde{\Phi}_m \right\|^2 & \leq \frac{C}{\omega(k)^2} \left[\|\Delta_{-h}G_\lambda\tilde{\Phi}_m\|^2 + \|\nabla_x \Delta_{-h}G_\lambda|\tilde{\Phi}_m\|^2 \right. \\ & \quad \left. + \omega(k)^2 \|\Delta_{-h}(\omega w_\lambda)\tilde{\Phi}_m\|^2 + \|a_\lambda(k-h)\tilde{\Phi}_m\|^2 |\Delta_h\omega(k-h)|^2 \right]. \end{aligned} \quad (145)$$

Combining this with (136) we get

$$\begin{aligned} & \|(\Delta_{-h}a_\lambda)(k)\tilde{\Phi}_m\|^2 \\ & \leq \frac{C}{\omega(k)^2} \left[\|\Delta_{-h}G_\lambda\tilde{\Phi}_m\|^2 + \|\nabla_x\Delta_{-h}G_\lambda|\tilde{\Phi}_m\|^2 + \omega(k)^2\|\Delta_{-h}(\omega w_\lambda)\tilde{\Phi}_m\|^2 \right] \\ & \quad + \frac{C}{\omega(k)^2|k-h|} \| |x|\tilde{\Phi}_m \|^2 \chi_\Lambda(k-h) |\Delta_h\omega(k-h)|^2. \end{aligned} \quad (146)$$

The polarization vectors defined in (59), (60), are differentiable away from the 3-axis. The same straightforward estimates as in Sect. 6 lead to

$$\begin{aligned} & \|(\Delta_{-h}a_\lambda)(k)\tilde{\Phi}_m\|^2 \\ & \leq C \left[\frac{1}{|k|(k_1^2 + k_2^2)} + \frac{1}{|k-h|((k_1-h_1)^2 + (k_2-h_2)^2)} \right] |h|^2 \| |x|\tilde{\Phi}_m \|^2 \end{aligned} \quad (147)$$

which hold for all $|k| < \Lambda$ and small $|h|$ with a constant C that is independent of m .

Next, we observe that for $k \neq 0$ fixed, there exist a sequence of h values, say h_l , tending to zero, so that $h^{-1}(\Delta_{-he_j}a_\lambda)(k)\tilde{\Phi}_m$ converges weakly to some element $v_j(k)$ which satisfies the estimate

$$\|v_j(k)\|^2 \leq C \frac{1}{|k|(k_1^2 + k_2^2)} \| |x|\tilde{\Phi}_m \|^2. \quad (148)$$

Here e_j is the j -th canonical basis vector. Next we identify $v_j(k)$ as the weak derivative of $a_\lambda(k)\tilde{\Phi}_m$. This weak derivative, by definition, can be computed via the expression

$$-(\Psi, a(\partial_j\phi)\tilde{\Phi}_m) \quad (149)$$

where Ψ is any state in $\mathcal{H}$ and ϕ is any test function in $C_c^\infty(\mathbb{R}^3)$. Clearly the above expression equals

$$\lim_{h \rightarrow 0} \int (\Psi, (\Delta_{-he_j}a_\lambda)(k)\tilde{\Phi}_m)\phi(k)dk. \quad (150)$$

But along the sequence h_l

$$\lim_{l \rightarrow \infty} \int (\Psi, \Delta_{-h_l e_j}a_\lambda(k)\tilde{\Phi}_m)\phi(k)dk = \int (\Psi, v_j(k))\phi(k)dk, \quad (151)$$

which identifies $v_j(k)$ as the (negative) weak derivative of $a_\lambda(k)\tilde{\Phi}_m$. $\square$

References

594

1. S. Agmon, Lectures on exponential decay of solutions of second order elliptic equations: Bounds on eigenfunctions of N-body Schrödinger operators, *Mathematical Notes* **29**, Princeton University Press (1982)
2. A. Arai, Rigorous theory of spectra and radiation for a model in quantum electrodynamics, *J. Math. Phys.* **24**, 1896–1910 (1983)
3. A. Arai, M. Hirokawa, On the existence and uniqueness of ground states of a generalized spin-boson model, *J. Funct. Anal.* **151**, 455–503 (1997)
4. A. Arai, M. Hirokawa, Ground states of a general class of quantum field Hamiltonians, *Rev. Math. Phys.* **12**, 1085–1135 (2000), mp_arc 99–179 (1999)
5. A. Arai, M. Hirokawa, F. Hiroshima, On the absence of eigenvectors of Hamiltonians in a class of massless quantum field models without infrared cutoff, *J. Funct. Anal.* **168**, 470–497 (1999)
6. V. Bach, J. Fröhlich, I.M. Sigal, Mathematical theory of nonrelativistic matter and radiation, *Lett. Math. Phys.* **34**, 183–201 (1995)
7. V. Bach, J. Fröhlich, I.M. Sigal, Quantum electrodynamics of confined non-relativistic particles, *Adv. Math.* **137**, 299–395 (1998)
8. V. Bach, J. Fröhlich, I.M. Sigal, Spectral analysis for systems of atoms and molecules coupled to the quantized radiation field, *Commun. Math. Phys.* **207**, 249–290 (1999)
9. H. Bethe, The electromagnetic shift of energy levels, *Phys. Rev.* **72**, 339–342 (1947)
10. J. Combes, L. Thomas, Asymptotic behavior of eigenfunctions for multiparticle Schrödinger operators, *Commun. Math. Phys.* **34**, 251–270 (1973)
11. J. Dereziński, C. Gérard, Asymptotic completeness in quantum field theory. Massive Pauli-Fierz Hamiltonians, *Rev. Math. Phys.* **11**, 383–450 (1999)
12. M. Dresden, H.A. Kramers, *Between tradition and revolution*, Springer Verlag (1987)
13. J. Fröhlich, On the infrared problem in a model of scalar electrons and massless scalar bosons, *Ann. Inst. H. Poincaré* **19**, 1–103 (1973)
14. J. Fröhlich, Existence of dressed one-electron states in a class of persistent models, *Fortschritte Phys.* **22**, 159–198 (1974)
15. C. Gérard, On the existence of ground states for massless Pauli-Fierz Hamiltonians, *Ann. Henri Poincaré* **1**, 443–459 (2000), mp_arc 99–158 (1999)
16. M. Hirokawa, Remarks on the ground state energy of the spin-boson model, An application of the Wigner-Weisskopf model, *Rev. Math. Phys.* **13**, 221–251 (2001), mp_arc 00–239 (2000)
17. F. Hiroshima, Ground states of a model in nonrelativistic quantum electrodynamics I and II, *J. Math. Phys.* **40**, 6209–6222 (1999), **41**, 661–674 (2000)
18. F. Hiroshima, The self-adjointness and relative bound of the Pauli-Fierz Hamiltonian in quantum electrodynamics for arbitrary coupling constants, preprint (October, 2000)
19. F. Hiroshima, H. Spohn, Enhanced binding through coupling to a quantum field, *Mathematical Physics Preprint Archive*, mp_arc 01–39 (2001)
20. W. Hunziker, I.M. Sigal, The general theory of N-body quantum systems, in: *Mathematical quantum theory. II. Schrödinger operators* (Vancouver, BC, 1993), 35–72, CRM Proc. Lecture Notes, **8**, Amer. Math. Soc., Providence, RI, 1995
21. E.H. Lieb, M. Loss, Self-Energy of Electrons in Non-perturbative QED, in: *Differential Equations and Mathematical Physics*, University of Alabama, Birmingham, 1999, R. Weikard, G. Weinstein, eds., 255–269, Internat. Press (1999). arXiv math-ph/9908020, mp_arc 99–305
22. E.H. Lieb, M. Loss, *Analysis*, Graduate Studies in Mathematics, American Mathematical Society, 1997
23. A. O'Connor, Exponential decay of bound-state wave functions, *Commun. Math. Phys.* **32**, 319–340 (1973)
24. W. Pauli, M. Fierz, Zur Theorie der Emission langwelliger Lichtquanten, *Nuovo Cimento* **15**, 167–188 (1938)

595

25. M. Reed, B. Simon, *Methods of modern mathematical physics*, vol 4, Theorem XIII.39, Academic Press (1978)
26. H. Spohn, Asymptotic completeness for Rayleigh scattering, *J. Math. Phys.* **38**, 2281–2296 (1997)
27. H. Spohn, Ground state(s) of the spin-boson Hamiltonian, *Commun. Math. Phys.* **123**, 277–304 (1989)
28. H. Spohn, Ground state of a quantum particle coupled to a scalar Bose field, *Lett. Math. Phys.* **44**, 9–16 (1998)

© 2003 by the authors
Adv. Theor. Math. Phys. 7 (2003) 667–710

Existence of Atoms and Molecules in Non-Relativistic Quantum Electrodynamics

Elliott H. Lieb¹ and Michael Loss²

¹Departments of Physics and Mathematics, Jadwin Hall
Princeton University, P. O. Box 708, Princeton, NJ 08544
lieb@princeton.edu

²School of Mathematics, Georgia Tech, Atlanta, GA 30332
loss@math.gatech.edu

Abstract

We show that the Hamiltonian describing N nonrelativistic electrons with spin, interacting with the quantized radiation field and several fixed nuclei with total charge Z , has a ground state when $N < Z + 1$. The result holds for any value of the fine structure constant α and for any value of the ultraviolet cutoff Λ on the radiation field. There is no infrared cutoff. The basic mathematical ingredient in our proof is a novel localization of the electromagnetic field in such a way that the errors in the energy are of smaller order than $1/L$, where L is the localization radius.

e-print archive: <http://lanl.arXiv.org/abs/math-ph/0307046>

¹Work partially supported by U.S. National Science Foundation grant PHY 01-39984.

²Work partially supported by U.S. National Science Foundation grant DMS 03-00349.
This paper may be reproduced, in its entirety, for non-commercial purposes.

1 Introduction

The existence of atoms and molecules in the framework of the Schrödinger equation was proved by Zhislin [14] for fixed nuclei when $N < Z + 1$. That is to say, the bottom of the spectrum of the N -electron Hamiltonian is a genuine N -particle bound state that satisfies Schrödinger's equation with some energy E , for each choice of the locations of the nuclei. (Here N is the number of electrons, each of charge $-e$, and Ze is the total charge of one or more fixed, positively charged nuclei.) The main physical result of the present paper is the proof of the same thing when account is taken of the ever-present quantized electromagnetic field. The interaction of this field with the electrons (but not the field itself) necessarily has an ultraviolet cutoff $|k| \leq \Lambda$ (in order to have finite quantities), but we emphasize that no infrared cutoff is used here.

If the fine structure constant $\alpha = e^2/\hbar c$ and Λ are small enough, the result follows from [1], but our result holds for *all values* of these parameters. In a recent paper Barbaroux, Chen and Vugalter [3] developed a new method that shows the existence of ground states for two-electron molecules with $2 < Z + 1$ (e.g., the Helium atom). Although they do not have to require that the perturbation is small when compared to the ionization energy as in [1], they have to impose restrictions on the various parameters since their works relies on the existence of the zero momentum ground state of the Hamiltonian of an electron interacting only with the radiation field. This has been established in [4] but only for sufficiently small coupling constants. The method of [3] is different from ours.

Our work (and [3]) relies on earlier work with Griesemer [7] where it was shown that a ground state exists provided a “binding condition” is satisfied, and it is this condition that is proved in [3] for the restricted $N = 2$ case and for the general case here for $N < Z + 1$. If $E^V(N)$ denotes the bottom of the spectrum of the Hamiltonian $H^V(N)$, which includes the Coulomb attraction of the electrons to the fixed nuclei of various positive charges $Z_1e, \dots, Z_Ke$ with $Z = \sum Z_j$, and if $E^0(N)$ denotes the bottom of the spectrum of $H^0(N)$ — the “free-electron” Hamiltonian in which there are no nuclei, but the electron-electron Coulomb repulsion is included — then the *binding condition* is

$$E^V(N) < \min \{ E^V(N') + E^0(N - N') : 0 \leq N' < N \} . \quad (1.1)$$

This binding condition, incidentally, is the same condition that Zhislin derived for the Schrödinger equation without the quantized electromagnetic field, and which he verified for $N < Z + 1$.

The inclusion of the quantized electromagnetic field presents two main difficulties. One is that if the bottom of the spectrum contains an eigenvalue it is *not* an isolated eigenvalue, as it was in [14]. Rather, the bottom of the spectrum is always the bottom of the essential spectrum because one can create arbitrarily many, arbitrarily soft photons. It is not easy to find an eigenvalue when it lies in the continuum. This problem was solved in [7] under condition (1.1).

The second main problem, which complicates the proof of (1.1), comes from the fact that each electron carries a virtual cloud of photons. This cloud may have substantial energy and when two electrons are near each other (whether bound or not) the interference of the photon clouds must be taken into account. In general, this is a highly non-perturbative effect. Our way around this difficulty is to prove that the photon clouds can be localized (i.e., effectively eliminated outside a ball of radius L surrounding the electron or the atom) in such a way that the error induced in the energy of the cloud is smaller than $L^{-(1+\varepsilon)}$, and thus the direct Coulomb interaction, which goes as L^{-1} , is dominant — as it was in the original paper [14]. A closely related effect is that even in the absence of an external potential electrons interact with each other. In such a case their dynamics is governed by H^0 , which contains the electron-electron Coulomb repulsion. Nevertheless, it is not inconceivable that the quantized field, which interacts simultaneously with all the electrons, might cause binding among the “free” electrons. While this is unlikely it has never been disproved and we must not assume in our proof that $E^0(N) = NE^0(1)$.

We are grateful to the anonymous referee who made many valuable suggestions and who helped us understand some conceptual matters about photons.

2 Basic Definitions and Concepts

The Hamiltonian under consideration, in appropriate units, is the Pauli-Fierz Hamiltonian and is given by

$$H^V(N) = \sum_{j=1}^N \left[(p_j + \sqrt{\alpha} A(x_j))^2 + \frac{g}{2} \sqrt{\alpha} \sigma_j \cdot B(x_j) + V(x_j) \right] + \alpha \sum_{i < j} \frac{1}{|x_i - x_j|} + H_f. \quad (2.1)$$

Here, g is some constant (close to 2, physically) and the vector σ_j is the set of three Pauli spin matrices for electron j . (Owing to the ultraviolet cutoff there is no restriction to $|g| \leq 2$, as there would be without a cutoff [12].) The operator p_j denotes $-i\nabla$ acting on the coordinate of the j -th electron.

The potential V is the potential of $K \geq 1$ nuclei with positive charges $Z_1, \dots, Z_K$ and locations $R_1, \dots, R_K \in \mathbb{R}^3$.

$$V(x) = - \sum_{j=1}^K Z_j |x - R_j|^{-1}. \quad (2.2)$$

Remark: The truth of our main theorem (3.1) — and its proof — does not require that $V(x)$ be given by (2.2). In addition to the general condition [7, eq. (5)], we need only the condition that there is some radius ρ such that $\langle V(x) \rangle \leq -Z/|x|$ for $|x| > \rho$, where $\langle \cdot \rangle$ denotes spherical average. Similarly, the repulsion $|x_i - x_j|^{-1}$ can be replaced by $W(x_i - x_j)$ provided $\langle W(x) \rangle \leq 1/|x|$ for all $|x| > \rho$.

The free Hamiltonian $H^0(N)$ is similar to $H^V(N)$, but without the attraction to the nuclei, i.e.,

$$H^0(N) = \sum_{i=1}^N \left[(p_i + \sqrt{\alpha} A(x_i))^2 + \frac{g}{2} \sqrt{\alpha} \sigma_i \cdot B(x_i) \right] + \alpha \sum_{i < j} \frac{1}{|x_i - x_j|} + H_f. \quad (2.3)$$

Note that the Coulomb repulsion among the electrons is included. The reason for including the Coulomb repulsion is (as stated above) that we do not know whether the electrons bind to each other through the interaction with the electromagnetic field, i.e., the electrons may not separate in the lowest energy state (if there is one).

The (ultraviolet cutoff) magnetic vector potential is defined by

$$A(x) = \frac{1}{2\pi} \sum_{\lambda=1}^2 \int \frac{\varepsilon_{\lambda}(k)}{\sqrt{|k|}} \hat{\chi}_{\Lambda}(k) \left(\hat{a}_{\lambda}(k) e^{ik \cdot x} + \hat{a}_{\lambda}^*(k) e^{-ik \cdot x} \right) dk, \quad (2.4)$$

where the function $\hat{\chi}_{\Lambda}$ is a smooth, radial function in k space, that vanishes outside the ball whose radius is the ultraviolet cutoff Λ . We denoted the creation and destruction operators of photons of momentum k and polarization λ by $\hat{a}_{\lambda}(k)$ and $\hat{a}_{\lambda}^*(k)$. This unusual notation is used since we shall later introduce the creation and destruction operators in configuration space, $a_{\lambda}(y)$ and $a_{\lambda}^*(y)$, which act on the Fourier transformed functions in Fock space.

The magnetic field is $B(x) = \text{curl} A(x)$. The operators $\widehat{a}_\lambda, \widehat{a}_\lambda^*$ satisfy the usual commutation relations

$$[\widehat{a}_\lambda(k), \widehat{a}_\nu^*(q)] = \delta(k - q)\delta_{\lambda,\nu}, \quad [\widehat{a}_\lambda(k), \widehat{a}_\nu(q)] = 0, \quad \text{etc} \quad (2.5)$$

and the vectors $\varepsilon_\lambda(k)$ are the two possible orthonormal polarization vectors perpendicular to k and to each other.

The vectors $\varepsilon_\lambda(k)$ have to be discontinuous functions of k on every sphere of fixed $|k|$ -value because it is not possible to “comb the hair on a sphere”. However, the only physical quantity,

$$\sum_{\lambda=1}^2 \varepsilon_\lambda^i(k) \varepsilon_\lambda^j(k) = \delta_{i,j} - \frac{k_i k_j}{|k|^2}, \quad (2.6)$$

is discontinuous only at the point $k = 0$. For the rest of this paper we choose the polarizations vectors to be

$$\begin{aligned} \varepsilon_1(k) &= \frac{(k_2, -k_1, 0)}{\sqrt{k_1^2 + k_2^2}}, \\ \varepsilon_2(k) &= \frac{k}{|k|} \wedge \varepsilon_1(k). \end{aligned} \quad (2.7)$$

Let us emphasize here that some smoothness of the function $\widehat{\chi}_\Lambda$ is essential for our arguments since this guarantees that the coupling functions

$$h_\lambda^i(y) = \frac{1}{2\pi} \int \frac{\widehat{\chi}_\Lambda(k)}{\sqrt{|k|}} \varepsilon_\lambda^i(k) e^{-ik \cdot y} dk \quad (2.8)$$

has a suitable decay as $|y| \rightarrow \infty$. If we did not have the discontinuous function $\varepsilon_\lambda^i(k)$ in (2.8) then $h(y)$ would decay as $|y|^{-5/2}$ as $|y| \rightarrow \infty$. (Proof: $|k|^{-1/2}$ is the Fourier transform of $|y|^{-5/2}$ in the sense of distributions [10, Theorem 5.9]. The Fourier transform of $\widehat{\chi}_\Lambda$ is real analytic and decays faster than any inverse power of $|y|$. Hence, the convolution of χ with $|y|^{-5/2}$ decays like $|y|^{-5/2}$. With a sharp cutoff it would decay only like $|y|^{-2}$ which turns out to be insufficient for a good localization of the photon states.

This analysis of h shows that we have to be circumspect about the choice of the polarization vectors. Their discontinuity will spoil the $|y|^{-5/2}$ decay, but it is important to get better decay than $|y|^{-2}$. In Lemma B.1 of Appendix B it is shown that with our choice (2.7) of the polarization vectors the coupling functions have sufficient decay in the sense that $\int |y|^{2\gamma} |h_\lambda^i(y)|^2 dy$ is finite for all $\gamma < 1$. Thus, in an average sense, the coupling functions decay

almost as fast as $|y|^{-5/2}$. We made no attempt to optimize the choice of the polarization vectors.

While polarization is physically measurable, the polarization vectors are not. They are merely a basis. It is odd, therefore, that their mathematical definition plays a role in the spatial localization of the photon field that we shall construct, and which is central to our proof of the binding condition. It would be better to start with a formalism that contains only “divergence-free” vector fields as the dynamical variables instead of trying to define them with the aid of unphysical polarization vectors. In particular, the Fock space would be built over the L^2 -space of divergence-free vector fields instead of $L^2 \otimes \mathbb{C}^2$. We shall not explore this here, but we mention that the localization of a divergence-free vector field, which preserves the divergence-free property is also a subtle matter.

The field energy, H_f , sometimes called $d\Gamma(\omega)$, is given by

$$H_f = \sum_{\lambda=1,2} \int_{\mathbb{R}^3} |k| \hat{a}_\lambda^*(k) \hat{a}_\lambda(k) dk \quad (2.9)$$

There is no cutoff in H_f . The energy of a photon is $|k|$.

Another unbounded operator of interest is the number operator

$$\mathcal{N} = \sum_{\lambda=1,2} \int_{\mathbb{R}^3} \hat{a}_\lambda^*(k) \hat{a}_\lambda(k) dk . \quad (2.10)$$

The physical Hilbert space for this system is given by

$$\mathcal{H}(N) = \wedge^N L^2(\mathbb{R}^3; \mathbb{C}^2) \otimes \mathcal{F} \quad (2.11)$$

where the wedge indicates that the electron wave functions are antisymmetric under the exchange of the particle labels. Thus, the functions in the space $\mathcal{H}(N)$ obey the Pauli exclusion principle. The photon Fock space is $\mathcal{F}$. We denote the inner product of two states Ψ and Φ in the space $\mathcal{H}(N)$ or in Fock space alone by

$$(\Psi, \Phi) \quad \text{and} \quad \langle \Psi, \Phi \rangle , \quad (2.12)$$

respectively. If Ψ and Φ are in $\mathcal{H}(N)$ then $\langle \Psi, \Phi \rangle$ makes sense and defines a summable function of $x_1, s_1, \dots, x_N, s_N$, where x_j, s_j are the space-spin variables of the j -th electron.

It is desirable that the above Hamiltonians be selfadjoint on certain domains and this has been worked out, e.g., in [9]. In this paper we will always

be talking about the Friedrichs extension of the symmetric operators $H^\#(N)$ (where $\#$ is 0 or V). The form domain will consist of all states for which each term in the operators has a finite expectation value. Accordingly, we define the ground state energy $E^\#(N)$ for the Hamiltonian $H^\#(N)$ by

$$E^\#(N) = \inf \left\{ \left(\Psi, H^\#(N) \Psi \right) : \Psi \in \mathcal{H}(N), \|\Psi\| = 1 \right\}. \quad (2.13)$$

The numbers $E^\#(N)$ are finite. This follows from Lemma A.4 in [7] together with the fact that the Coulomb potential is form bounded with respect to $p^2 = -\Delta$.

A few remarks concerning the Fock space $\mathcal{F}$ are in order. It is built over the space $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$; the second factor takes into account the polarizations. Let $\{f_i\}$, $i = 1, 2, \dots$, be an orthonormal basis for $L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$. Then, vectors of the form

$$|i_1, m_1; \dots; i_n, m_n\rangle = \frac{1}{\sqrt{m_1! \dots m_n!}} a^*(f_{i_1})^{m_1} \dots a^*(f_{i_n})^{m_n} |0\rangle, \quad (2.14)$$

constitute an orthonormal basis for $\mathcal{F}$, the occupation number basis. In (2.14) n is an arbitrary nonnegative integer (with $n = 0$ denoting the vacuum vector $|0\rangle$), the indices $i_1, \dots, i_n$ are all different, the m_i are all positive integers, $a^*(f)$ is an abbreviation for $\sum_\lambda a_\lambda^*(f_\lambda)$ and $f_\lambda = f(k, \lambda)$ is a function in L^2 . Thus, any state $\Phi \in \mathcal{F}$ can be uniquely written as

$$\Phi = \sum_{n \geq 0} \sum_{i_1 < i_2 < \dots < i_n} \sum_{m_1, \dots, m_n} \phi_{i_1, m_1; \dots; i_n, m_n} |i_1, m_1; \dots; i_n, m_n\rangle, \quad (2.15)$$

where the $n = 0$ term in (2.15) is just $\phi_0 |0\rangle$ with $\phi_0 \in \mathbb{C}$. The inner product is given by

$$\langle \Phi, \Phi \rangle = \sum_{n \geq 0} \sum_{i_1 < i_2 < \dots < i_n} \sum_{m_{i_1}, \dots, m_{i_n}} |\phi_{i_1, m_1; \dots; i_n, m_n}|^2. \quad (2.16)$$

This representation has the advantage that the symmetry in the photon variables is automatically taken care of. It is particularly useful when dealing with product states. Consider a state Φ whose photons are all **localized** in a closed region $\mathcal{Y} \subset \mathbb{R}^3$. This means that all the $f_i(y, \lambda)$ appearing in (2.14) and in (2.15) vanish if $y \notin \mathcal{Y}$. Likewise, consider a state Ψ whose photons are all localized in a closed region $\mathcal{Z} \subset \mathbb{R}^3$ which is disjoint from $\mathcal{Y}$. Pick an orthonormal basis $\{f_k\}$ in $L^2(\mathcal{Y}) \otimes \mathbb{C}^2$ and an orthonormal basis $\{g_\ell\}$ in $L^2(\mathcal{Z}) \otimes \mathbb{C}^2$. Clearly, the two algebras of creation and annihilation operators generated by $a^\#(f_k)$ and $a^\#(g_\ell)$ commute. If

$$\Phi = \sum_{n \geq 0} \sum_{i_1 < i_2 < \dots < i_n} \sum_{p_1, \dots, p_n} \phi_{i_1, p_1; \dots; i_n, p_n} |i_1, p_1; \dots; i_n, p_n\rangle_{\mathcal{Y}} \quad (2.17)$$

and

$$\Psi = \sum_{n \geq 0} \sum_{j_1 < j_2 < \dots < j_n} \sum_{q_1, \dots, q_n} \psi_{j_1, q_1; \dots; j_n, q_n} |j_1, q_1; \dots; j_n, q_n\rangle_{\mathcal{Z}} \quad (2.18)$$

then we define the product state Ξ by

$$\Xi = \sum \phi_{i_1, p_1; \dots; i_m, p_m} \psi_{j_1, q_1; \dots; j_k, q_k} |i_1, p_1; \dots; i_m, p_m\rangle_{\mathcal{Y}} \otimes |j_1, q_1; \dots; j_k, q_k\rangle_{\mathcal{Z}} \quad (2.19)$$

where

$$|i_1, p_1; \dots; i_m, p_m\rangle_{\mathcal{Y}} \otimes |j_1, q_1; \dots; j_k, q_k\rangle_{\mathcal{Z}} = \frac{1}{\sqrt{p_1! \dots p_m!} \sqrt{q_1! \dots q_k!}} a^*(f_{i_1})^{p_1} \dots a^*(f_{i_m})^{p_m} a^*(g_{j_1})^{q_1} \dots a^*(g_{j_k})^{q_k} |0\rangle \quad (2.20)$$

By a simple calculation we find that

$$\langle \Xi, \Xi \rangle = \langle \Phi, \Phi \rangle \langle \Psi, \Psi \rangle. \quad (2.21)$$

Further, if f is a function supported in $\mathcal{Y}$, then

$$\langle \Xi, a^*(f) a(f) \Xi \rangle = \langle \Phi, a^*(f) a(f) \Phi \rangle \langle \Psi, \Psi \rangle. \quad (2.22)$$

Likewise, if f is supported in $\mathcal{Y}$ and g in $\mathcal{Z}$, then

$$\langle \Xi, a^*(f) a(g) \Xi \rangle = \langle \Phi, a^*(f) \Phi \rangle \langle \Psi, a(g) \Psi \rangle. \quad (2.23)$$

Quite generally, we have, for normal-ordered, bilinear expressions, the following formulas (in which β, γ denote linear forms in the annihilation operators a , and hence β^*, γ^* are linear forms in the creation operators):

$$\begin{aligned} \langle \Xi, \beta \gamma \Xi \rangle &= \langle \Psi, \beta \gamma \Psi \rangle \langle \Phi, \Phi \rangle + \langle \Psi, \Psi \rangle \langle \Phi, \beta \gamma \Phi \rangle \\ &\quad + \langle \Psi, \beta \Psi \rangle \langle \Phi, \gamma \Phi \rangle + \langle \Psi, \gamma \Psi \rangle \langle \Phi, \beta \Phi \rangle \\ \langle \Xi, \beta^* \gamma^* \Xi \rangle &= \langle \Psi, \beta^* \gamma^* \Psi \rangle \langle \Phi, \Phi \rangle + \langle \Psi, \Psi \rangle \langle \Phi, \beta^* \gamma^* \Phi \rangle \\ &\quad + \langle \Psi, \beta^* \Psi \rangle \langle \Phi, \gamma^* \Phi \rangle + \langle \Psi, \gamma^* \Psi \rangle \langle \Phi, \beta^* \Phi \rangle \\ \langle \Xi, \beta^* \gamma \Xi \rangle &= \langle \Psi, \beta^* \gamma \Psi \rangle \langle \Phi, \Phi \rangle + \langle \Psi, \Psi \rangle \langle \Phi, \beta^* \gamma \Phi \rangle \\ &\quad + \langle \Psi, \beta^* \Psi \rangle \langle \Phi, \gamma \Phi \rangle + \langle \Psi, \gamma \Psi \rangle \langle \Phi, \beta^* \Phi \rangle \end{aligned} \quad (2.24)$$

A formula of the type (2.24) does *not exist* for anti-normal-ordered products $\beta \gamma^*$. We shall have no need of such terms, however, because the only source of such terms is $A(x_i)^2$ in the electron kinetic energy. If we denote the part of (2.4) coming from $\widehat{a}_\lambda(k)$ by $\beta(x)$ and the remainder by $\beta^*(x)$ (see (2.29)) then

$$A(x)^2 = \beta(x)^2 + \beta^*(x)^2 + 2\beta^*(x)\beta(x) + C \quad (2.25)$$

where

$$C = \frac{1}{2\pi^2} \int \frac{|\widehat{\chi}_\Lambda(k)|^2}{|k|} dk. \quad (2.26)$$

Thus, apart from a fixed, finite number αNC , which is strictly proportional to N , (and which is, therefore, independent of any decomposition of the system into clusters) we can (and henceforth shall) replace $A(x_i)^2$ by the normal-ordered

$$: A(x_i)^2 := \beta(x_i)^2 + \beta^*(x_i)^2 + 2\beta^*(x_i)\beta(x_i). \quad (2.27)$$

Formulas (2.24) continue to hold for vectors Φ in the physical Hilbert space $\mathcal{H}(N)$, with the replacement of $\langle \cdot, \cdot \rangle$ by $(\cdot, \cdot)$. In this case, the coefficients, $\phi_{i_1, m_1; \dots; i_n, m_n}$, are (antisymmetric) functions of the electron space-spin coordinates, x_i, s_i .

It is convenient to introduce the operators given by

$$a_\lambda(y) = \frac{1}{(2\pi)^3} \int \widehat{a}_\lambda(k) e^{ik \cdot y} dk. \quad (2.28)$$

Then the vector potential can be written as

$$A^i(x) = \sum_{\lambda=1}^2 a_\lambda(h_\lambda^i(x - \cdot)) + a_\lambda^*(h_\lambda^i(x - \cdot)). \quad (2.29)$$

The action of the operators $a_\lambda(h_\lambda^i(x - \cdot))$ is given by

$$\begin{aligned} [a_\lambda(h_\lambda^i(x - \cdot))\Psi]_n(y_1, \lambda_1; \dots; y_n, \lambda_n) \\ = \sqrt{n+1} \int \overline{h_\lambda^i(x-y)} [\Psi]_{n+1}(y, \lambda; y_1, \lambda_1; \dots; y_n, \lambda_n) dy. \end{aligned} \quad (2.30)$$

A convenient expression for $a_\lambda(h_\lambda^i(x - \cdot))$ is the formula

$$a_\lambda(h_\lambda^i(x - \cdot)) = \int a_\lambda(y) \overline{h_\lambda^i(x-y)} dy. \quad (2.31)$$

The number operator and the field energy can be expressed in terms of the operators $a_\lambda(x)$ by

$$(\Phi, \mathcal{N}\Phi) = (2\pi)^3 \sum_{\lambda=1}^2 \int \|a_\lambda(x)\Phi\|^2 dx, \quad (2.32)$$

and

$$(\Phi, H_f\Phi) = (2\pi)^3 \sum_{\lambda=1}^2 \left(a_\lambda(\cdot)\Phi, \sqrt{-\Delta} a_\lambda(\cdot)\Phi \right) \quad (2.33)$$

which by eq. 7.12(4) in [10] can be rewritten as

$$4\pi \sum_{\lambda=1}^2 \int \frac{\|a_\lambda(x)\Phi - a_\lambda(y)\Phi\|^2}{|x-y|^4} dx dy. \quad (2.34)$$

By the previous considerations we have for the product state Ξ that

$$\begin{aligned} \sum_{\lambda} (a_\lambda(x)\Xi, a_\lambda(y)\Xi) = \\ \sum_{\lambda} (a_\lambda(x)\Phi, a_\lambda(y)\Phi) (\Psi, \Psi) + (\Phi, \Phi) \sum_i (a_\lambda(x)\Psi, a_\lambda(y)\Psi) \\ + \sum_{\lambda} (a_\lambda(x)\Phi, \Phi) (\Psi, a_\lambda(y)\Psi) + \sum_{\lambda} (\Phi, a_\lambda(y)\Phi) (a_\lambda(x)\Psi, \Psi), \end{aligned} \quad (2.35)$$

and hence we obtain for the field energy of Ξ the expression

$$\begin{aligned} (\Xi, H_f\Xi) = (\Phi, H_f\Phi) (\Psi, \Psi) + (\Phi, \Phi) (\Psi, H_f\Psi) \\ - 8\pi \sum_{\lambda} \Re \left[\int \frac{(a_\lambda(x)\Phi, \Phi) (\Psi, a_\lambda(y)\Psi) + (\Phi, a_\lambda(y)\Phi) (a_\lambda(x)\Psi, \Psi)}{|x-y|^4} \right] dx dy. \end{aligned} \quad (2.36)$$

The x integration in the first term of the last integral runs over the set $\mathcal{Y}$ while the y integration runs over the set $\mathcal{Z}$ and similarly in the second term the x integration runs over the set $\mathcal{Z}$ while the y integration runs over the set $\mathcal{Y}$. Hence the last expression is well defined as long as the distance between the sets $\mathcal{Y}$ and $\mathcal{Z}$ is positive. This term expresses the fact that the field energy is a nonlocal operator and this nonlocality is one of the main obstacles to be overcome.

In general, the states Φ and Ψ will depend on the position and spin variables of the various electrons and hence the product state (2.19) has to be antisymmetrized over the electron labels. It is straightforward to check

that the expression (2.36) continues to hold also for such states. (When different groups of electrons are involved an antisymmetrization is required, however, as discussed in (3.8).)

We need one more concept before stating our main theorem. It will be necessary to localize both the electrons and the photon field. As far as the electrons are concerned it is useful to define what we mean by a **symmetrized product of n domains** in $\mathbb{R}^3$. If $B_1, \dots, B_n$ are n domains (open sets) in $\mathbb{R}^3$ then the symmetrized product, Ω , is a domain in $(\mathbb{R}^3)^n$ given by

$$\Omega(B_1, \dots, B_n) = \bigcup_{\pi \in S_n} B_{\pi 1} \times B_{\pi 2} \times \cdots \times B_{\pi n}, \quad (2.37)$$

where S_n is the group of permutations of n labels. It might be useful to illustrate this when $n = 2$. Then we have $B_1 \subset \mathbb{R}^3$, $B_2 \subset \mathbb{R}^3$ and $\Omega(B_1, B_2) = (B_1 \times B_2) \cup (B_2 \times B_1)$. This is different from $(B_1 \cup B_2) \times (B_2 \cup B_1)$. Physically it means there is one particle in each domain B_i , but the label of the particle is indeterminate. If the domains overlap there may be several particles in one domain, of course.

3 The Main Theorem

The following is our main theorem. The proof given in this section uses several inequalities derived later on in this paper, but we present the proof now in order to make the main ideas clear without too many technicalities.

THEOREM 3.1 (Binding in Atoms). *The strict inequality (1.1) holds for all $N < Z + 1$, all g , all α and all Λ . In particular this implies that there exists a normalized ground state $\Phi(N)$ in $\mathcal{H}(N)$ for the Hamiltonian $H^V(N)$, i.e., $(\Phi(N), H(N)\Phi(N)) = E^V(N)$, and it satisfies $H^V(N)\Phi(N) = E^V(N)\Phi(N)$.*

See the remark after eq. (2.2).

PROOF: Our proof has three main parts. The first is the construction of a good trial function for $N - N'$ (with $0 \leq N' < N$) localized, 'free' electrons and localized photons accompanying these localized electrons. The second part is the construction of a good trial function for N' localized electrons 'bound' to the given, fixed nuclei, together with localized photons. The third part consists in the construction of a trial function which is a product of these two functions and then showing that the energy is lowered (by a greater amount than the localization errors) because of a negative

Coulomb energy between the ‘bound’ system (consisting of electrons and nuclei) and the localized ‘free’ electrons. One difficulty in part 3 is that although the photons in the two regions are localized in separate regions, there is still a residual interaction between the two fields, given by the last term in (2.36), which has to be considered. This interaction comes from the fact that multiplication by $|k|$ in Fourier space is a nonlocal operation in position space.

The general argument proceeds by induction. We know from [7] that one electron binds. Assuming that the binding condition holds for M electrons, all $1 \leq M \leq N - 1$, we have to show that it holds for N electrons, i.e., $E^V(N) < \min\{E^V(N') + E^0(N - N') : 0 \leq N' < N\}$. Using [7, Theorem 2.1] we may assume that the Hamiltonian $H^V(M)$ has a ground state for all $1 \leq M \leq N - 1$. By the second part of [7, Theorem 3.1] we know that $E^V(N) < E^0(N)$ for all N , since $Z > 0$ and the attractive Coulomb potential is strictly negative.

Part1. From now on we set

$$n = N - N'.$$

Given $0 \leq N' < N$ we shall construct a normalized state $\Phi(n)$ for the free electron Hamiltonian $H^0(n)$ with the property that the n electrons are localized in a symmetrized product Ω of some balls of radius R_0 while the field is localized in balls with the same center but with radius $L > 2R_0$. The construction of $\Phi(n)$ is done in Theorem 4.3. It lies in the physical Hilbert space $\mathcal{H}(n)$ and has an energy given by

$$\frac{(\Phi(n), H^0(n)\Phi(n))}{(\Phi(n), \Phi(n))} \leq E^0(n) + \frac{Cn}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L^\gamma} \right) (1 + |\log(\Lambda R_0)|) + \frac{2\pi^2 n^2}{R_0^2}, \quad (3.1)$$

for any $\gamma < 1$, where C is some constant independent of L and R_0 . (It does depend on γ and on n , but n is bounded by N).

In (3.1) the term $(Cn/(L - 2R_0)^\gamma) (R_0/L^\gamma) (1 + |\log(\Lambda R_0)|)$ comes from the energy needed to localize the field in n balls of radius L . The last term comes from the kinetic energy needed to localize n electrons in the n balls of radius R_0 (Lemma 4.1).

Part2. According to the induction assumption at the beginning of this proof we may assume that $1 \leq N' \leq N - 1$ and that the Hamiltonian $H^V(N')$ of the bound electrons has a normalized ground state $\Gamma(N')$. By [7, Lemma 6.2] we know that this ground state is exponentially localized in the electron variables, i.e., if we denote by $|X|$ the quantity $\sum_{i=1}^{N'} |x_i|$ then

$$\|e^{\beta|X|} \Gamma(N')\|^2 \leq C_\beta \quad (3.2)$$

for any $\beta^2 < \min\{E^V(N' - m) + E^0(m) : 0 < m \leq N'\} - E^V(N')$.

(Note: An error in the proof of this exponential localization in [7, Lemma 6.2] was discovered by J-M. Barbaroux and the necessary correction was published in [6]. We are grateful to Prof. Barbaroux for pointing out this mistake to us.)

Although it is not necessary to do so, we (strictly) localize $\Gamma(N')$ so that all the electrons are in a common ball of radius R_0 . Following that, we localize the photon field in a larger ball of radius $L > R_0$. The field localization is essential. The electron localization is not since it would be possible to use only the exponential decay of $\Gamma(N')$. The localization is done as follows.

Let $\chi \leq 1$ be a smooth cutoff function with support in the unit ball centered at the origin and $\chi(x) = 1$ for $|x| < 1/2$. Let $\Theta = \prod_{i=1}^{N'} \chi(x_i/R_0)$ and $\tilde{\Gamma}(N') = \Theta\Gamma(N')$. Since $\Gamma(N')$ is a ground state, and hence satisfies the Schrödinger equation, we can deduce that the increase in energy due to the cutoff is bounded as follows.

$$\left(\tilde{\Gamma}(N'), H^V(N')\tilde{\Gamma}(N')\right) \leq E^V(N') \left(\tilde{\Gamma}(N'), \tilde{\Gamma}(N')\right) + N' \frac{C_\beta \exp(-\beta R_0)}{R_0^2}. \quad (3.3)$$

Inequality (3.3) follows from (3.2) and integration by parts as follows. With $\langle, \rangle$ denoting inner product in Fock space and dX denoting integration over the space-spin variables, we have

$$\begin{aligned} \int \langle \Theta\Gamma(N'), \sum_j (\nabla_j + iA(x_j))^2 \Theta\Gamma(N') \rangle dX = \\ \int \left(\sum_j \Theta \Delta_{x_j} \Theta \right) \langle \Gamma(N'), \Gamma(N') \rangle dX \\ + 2 \int \sum_j (\Theta \nabla_{x_j} \Theta) \cdot \langle \Gamma(N'), (\nabla_j + iA(x_j))\Gamma(N') \rangle dX \\ + \int \Theta^2 \langle \Gamma(N'), \sum_j (\nabla_j + iA(x_j))^2 \Gamma(N') \rangle dX. \end{aligned} \quad (3.4)$$

Since $(\nabla + iA(x))^2$ is a symmetric operator, the left side of (3.4) is real and, therefore, the right side must be real, too. The first term on the right side is real. The third term is also real because $\Gamma(N')$ satisfies the Schrödinger equation, and hence $\langle \Gamma(N'), \sum_j (\nabla_j + iA(x_j))^2 \Gamma(N') \rangle = \langle \Gamma(N'), (-E^V(N') + \text{real potentials})\Gamma(N') \rangle$, which is real. The middle term must, therefore, be

real too (when summed over all particles), and we can replace the integrand by its real part. This means that we can replace $2\langle \Gamma(N'), (\nabla_{x_j} + iA(x_j)), \Gamma(N') \rangle$ by $2\Re\langle \Gamma(N'), \nabla_{x_j} \Gamma(N') \rangle = \nabla_{x_j} \langle \Gamma(N'), \Gamma(N') \rangle$ since $\langle \Gamma(N'), iA(x_j) \Gamma(N') \rangle$ is imaginary (because $A(x_j)$ is symmetric).

Now, integrating by parts we can combine the second term with the first to yield $-\int (\nabla_{x_j} \Theta)^2 \langle \Gamma(N'), \Gamma(N') \rangle dx_j$. This is the error term (the last term leads to the principal term $E^V(N') (\tilde{\Gamma}(N'), \tilde{\Gamma}(N'))$). This error term can be bounded by replacing $\nabla \chi(x_j/R_0)$ by C/R_0 times the characteristic function of the annulus between $R_0/2$ and R_0 , for some constant C . But this characteristic function is bounded by $\exp(-\beta R_0/2) \exp(+\beta|x|)$. Inequality (3.3) then follows from the exponential decay (3.2).

Next one has to show that the error term in (3.3) is small when compared with $\|\tilde{\Gamma}(N')\|^2$. It follows from the exponential decay that

$$\|\tilde{\Gamma}(N')\|^2 \geq 1 - N' C_\beta e^{-\beta R_0} . \quad (3.5)$$

To see this note that $\chi(x_i/R_0)^2 \geq 1 - g_i$, where $g_i = 1$ if $|x_i| > R_0/2$ and $g_i = 0$ otherwise. Then $\Theta \geq \prod_i (1 - g_i) \geq 1 - \sum_i g_i$. But $g_i \leq \exp\{-\beta R_0\} \exp\{2\beta|x_i|\} \leq \exp\{-\beta R_0\} \exp\{2\beta|X|\}$. Therefore,

$$\begin{aligned} \|\tilde{\Gamma}(N')\|^2 &\geq \left(\Gamma(N'), \left(1 - \sum_i g_i \right) \Gamma(N') \right) \\ &\geq 1 - N' \exp\{-\beta R_0\} \left(\Gamma(N'), \exp\{-2\beta|X|\} \Gamma(N') \right) . \end{aligned}$$

Together, (3.3) and (3.5) imply (for $\exp\{\beta R_0\} > N' C_\beta$)

$$\begin{aligned} \frac{\left(\tilde{\Gamma}(N'), H^V(N') \tilde{\Gamma}(N') \right)}{\left(\tilde{\Gamma}(N'), \tilde{\Gamma}(N') \right)} &\leq E^V(N') + \frac{N'}{R_0^2} \frac{C_\beta}{\exp\{\beta R_0\} - N' C_\beta} \\ &\leq E^V(N') + \frac{2N'}{R_0^2} C_\beta e^{-\beta R_0} , \end{aligned} \quad (3.6)$$

where the last inequality holds provided that R_0 is chosen such that $\beta R_0 \geq \log(2N' C_\beta)$.

The next step is to localize the photons in the state $\tilde{\Gamma}(N')$ in a ball centered at the origin of radius $L > R_0$. This leads to a new state $\Psi(N')$ with

$$\frac{(\Psi(N'), H^V(N') \Psi(N'))}{(\Psi(N'), \Psi(N'))} \leq E^V(N') + \frac{N'}{R_0^2} \frac{C_\beta}{\exp\{\beta R_0\} - N' C_\beta} + \frac{CN'}{(L - 2R_0)^\gamma} \frac{R_0}{L^\gamma}, \quad (3.7)$$

for all $\gamma < 1$ and for R_0 and $L - 2R_0$ large enough.

The construction of this function Ψ is precisely the same as the photon field localization that led to the state $\Phi(n)$. It is carried out in Theorem 4.3.

Thus, $\Psi(N')$ is a state in which all the electrons are localized in a ball of radius R_0 and the photons are all localized in a ball of radius L . Moreover the localization errors are small and given in (3.7).

Part3. Now we put the pieces from Part 1 and Part 2 together and construct a trial function Ξ whose energy will be strictly below $E^V(N') + E^0(n)$.

As mentioned above, since $H^0(n)$ is translation invariant we can, by shifting, make sure that the photons in the state $\Psi(N')$ and the photons in the shifted state $\Phi(n)$, live in disjoint sets. This will be the case when the smallest distance of the centers of the balls $B_1, \dots, B_n$ with the center of the ball in which $\Psi(N')$ lives is greater than $2L$.

Now we can form the product state Ξ as indicated in (2.19). The state is symmetric in the photon variables by construction. It has to be antisymmetrized in the electron labels though, i.e., replace the products $\phi_{i_1, p_1; \dots; i_m, p_m} \psi_{j_1, q_1; \dots; j_k, q_k}$ in (2.19) by

$$c(N, N') \sum_{\pi \in S_N} (-1)^\pi \phi_{i_1, p_1; \dots; i_m, p_m}(z_{\pi(1)}, \dots, z_{\pi(N-N')}) \\ \times \psi_{j_1, q_1; \dots; j_k, q_k}(z_{\pi(N-N'+1)}, \dots, z_{\pi(N)}). \quad (3.8)$$

where π runs through all the permutations of N elements,

$$c(N, N') = \frac{1}{\sqrt{N!(N - N')!N'!}} \quad (3.9)$$

is the normalization and $z_j = (x_j, s_j)$, the position and spin of the j -th electron. The expression in (3.9) is calculated by noting first that ϕ and ψ are each antisymmetric in their electron coordinates and, second, that there is no overlap between ϕ and ψ because the x variables in the two functions have disjoint support. Informally speaking, the antisymmetrization in (3.8)

has no effect and could be dispensed with for all practical purposes since the operators ∇ and the potentials that we consider here are local operators. If we the Hamiltonian contained nonlocal operators, such as the ‘relativistic’ $\sqrt{-\Delta}$ then the antisymmetrization (3.8) would have a more profound effect — although (3.9) is still correct.

Next, we calculate (with $: :$ denoting normal ordering)

$$\begin{aligned} & (\Xi, H(N) \Xi) \\ &= \left(\Xi, \sum_{i=1}^N \left[(p_i + \sqrt{\alpha} A(x_i))^2 : + \frac{g}{2} \sqrt{\alpha} \sigma_i \cdot B(x_i) + V(x_i) \right] \Xi \right) \\ & \quad + \left(\Xi, \sum_{i < j} \frac{1}{|x_i - x_j|} \Xi \right) + (\Xi, H_f \Xi) \end{aligned} \quad (3.10)$$

in terms of the normalized $\Phi(n)$ and $\Psi(N')$. The field energy term has been explained previously in equation (2.36) and yields

$$\begin{aligned} & (\Xi, H_f \Xi) = (\Phi(n), H_f \Phi(n)) + (\Psi(N'), H_f \Psi(N')) \\ & \quad + 8\pi \sum_i \Re \int \frac{(a_\lambda(x) \Phi(n), \Phi(n)) (\Psi(N'), a_\lambda(y) \Psi(N'))}{|x - y|^4} dx dy \\ & \quad + 8\pi \sum_i \Re \int \frac{(\Phi(n), a_\lambda(y) \Phi(n)) (a_\lambda(x) \Psi(N'), \Psi(N'))}{|x - y|^4} dx dy. \end{aligned} \quad (3.11)$$

The Coulomb repulsion term is easily calculated to consist of three terms:

$$\left(\Phi(n), \sum_{N' < i < j \leq N} \frac{1}{|x_i - x_j|} \Phi(n) \right) (\Psi(N'), \Psi(N')) + \quad (3.12)$$

$$\left(\Psi(N'), \sum_{1 \leq i < j \leq N'} \frac{1}{|x_i - x_j|} \Psi(N') \right) (\Phi(n), \Phi(n)) + \quad (3.13)$$

$$\sum_{i=1, j=N'+1}^{N', N} \int \frac{\|\Psi(N')\|^2(x_1, \dots, x_{N'}) \|\Phi(n)\|^2(x_{N'+1}, \dots, x_N)}{|x_i - x_j|} d^N x \quad (3.14)$$

Here the norm signs indicate that the norm has been taken in Fock space and in the spin space.

The electron kinetic energy involves the calculation of terms of the form

$$\begin{aligned}
& (\Xi, a(f)^2 \Xi) \\
&= (\Psi(N'), a(f)^2 \Psi(N')) (\Phi(n), \Phi(n)) + (\Phi(n), a(f)^2 \Phi(n)) (\Psi(N'), \Psi(N')) \\
&\quad + 2\Re (\Psi(N'), a(f) \Psi(N')) (\Phi(n), a(f) \Phi(n)) . \quad (3.15)
\end{aligned}$$

Thus, we have that

$$\begin{aligned}
& \left(\Xi, \sum_{i=1}^N \left[: (p_i + \sqrt{\alpha} A(x_i))^2 : + \frac{g}{2} \sqrt{\alpha} \sigma_i \cdot B(x_i) \right] \Xi \right) = \\
& \sum_{i=1}^{N'} \left(\Psi(N'), \left[: (p_i + \sqrt{\alpha} A(x_i))^2 : + \frac{g}{2} \sqrt{\alpha} \sigma_i \cdot B(x_i) \right] \Psi(N') \right) (\Phi(n), \Phi(n)) \\
& + \sum_{i=N'+1}^N \left(\Phi(n), \left[: (p_i + \sqrt{\alpha} A(x_i))^2 : + \frac{g}{2} \sqrt{\alpha} \sigma_i \cdot B(x_i) \right] \Phi(n) \right) (\Psi(N'), \Psi(N')) \\
& + \alpha \sum_{i=1}^{N'} \int (\Phi(n), : A(x_i)^2 : \Phi(n)) \|\Psi(N')\|^2(x_1, \dots, x_{N'}) dx_1 \cdots dx_{N'} \\
& \quad (3.16)
\end{aligned}$$

$$\begin{aligned}
& + \alpha \sum_{i=N'+1}^N \int (\Psi(N'), : A(x_i)^2 : \Psi(N')) \|\Phi(n)\|^2(x_{N'+1}, \dots, x_N) dx_{N'+1} \cdots dx_N \\
& \quad (3.17)
\end{aligned}$$

$$\begin{aligned}
& + 2\alpha \sum_{i=1}^{N'} \int (\Phi(n), A(x_i) \Phi(n)) \cdot (\Psi(N'), p_i \Psi(N')) (x_1, \dots, x_{N'}) dx_1 \cdots dx_{N'} \\
& \quad (3.18)
\end{aligned}$$

$$\begin{aligned}
& + 2\alpha \sum_{i=N'+1}^N \int (\Psi(N'), A(x_i) \Psi(N')) \cdot \\
& \quad (\Phi(n), p_i \Phi(n)) (x_{N'+1}, \dots, x_N) dx_{N'+1} \cdots dx_N \quad (3.19)
\end{aligned}$$

$$\begin{aligned}
& + 2\alpha \sum_{i=1}^N \int \langle \Psi(N'), A(x_i) \Psi(N') \rangle \langle \Phi(n), A(x_i) \Phi(n) \rangle dx_1 \cdots dx_N \quad (3.20)
\end{aligned}$$

$$\begin{aligned}
& + \frac{g}{2} \sqrt{\alpha} \sum_{i=N'+1}^N \int (\Psi(N'), B(x_i) \Psi(N')) \cdot \\
& \quad (\Phi(n), \sigma_i \Phi(n)) (x_{N'+1}, \dots, x_N) dx_{N'+1} \cdots dx_N \quad (3.21)
\end{aligned}$$

$$\begin{aligned}
& + \frac{g}{2} \sqrt{\alpha} \sum_{i=1}^{N'} \int (\Phi(n), B(x_i) \Phi(n)) \cdot \\
& \quad (\Psi(N'), \sigma_i \Psi(N')) (x_1, \dots, x_{N'}) dx_1 \cdots dx_{N'} . \quad (3.22)
\end{aligned}$$

Finally, the last and most important term

$$\begin{aligned} \left(\Xi, \sum_{i=1}^N V(x_i) \Xi \right) &= \left(\Psi(N'), \sum_{i=1}^{N'} V(x_i) \Psi(N') \right) (\Phi(n), \Phi(n)) \\ &+ \left(\Phi(n), \sum_{i=N'+1}^N V(x_i) \Phi(n) \right) (\Psi(N'), \Psi(N')) \quad . \quad (3.23) \end{aligned}$$

Lemma 5.6 allows us to show that the terms (3.16, 3.17) are of order $L^{-2\gamma}$ for any $\gamma < 1$. This follows from the fact that in our trial function the electrons are localized in balls of radius R_0 , and so the distance D in Lemma 5.6 between any electron and the localized photon field of the subsystem to which the electron does *not* belong is at least $L - R_0$. Since we can easily choose L large and R_0 to be an arbitrarily chosen small constant times L we conclude, from Lemma 5.6, that these terms are of order $L^{-2\gamma}$ for any $\gamma < 1$.

The terms (3.14, 3.23) taken together are the terms that will give us binding. We shall show that, after averaging over rotations, the two terms add to $-(Z - N')/3L$, which is less than $-\text{pos. const.}/L$ according to our hypothesis.

‘*Averaging over rotations*’ means the following. We fix the state $\Psi(N')$ of the electrons bound to the nuclei and their field, which extends out a distance L from the origin. On the other hand, the state of the n unbound electrons was called $\Phi(n)$, but actually there are infinitely many states we could use. That is, we start with one $\Phi(n)$ and consider all rotations of it about the origin. The average Coulomb interaction (i.e., the average of (3.14) and (3.23)) is the same as if the bound electron state $\Psi(N')$, including the nuclei, was rotated about the origin. However, the average potential generated by the latter average over rotations at a point x would be exactly $(Z - N')/|x|$ *provided* $|x| > L$. This is Newton’s theorem [10, Theorem 9.7]. Therefore, there exists a rotation so that the Coulomb interactions (3.14) and (3.23) are as if the inner state were a point charge located at the origin and of strength $Z - N'$.

We now choose $\Phi(n)$ so that one of the balls of radius L in which the n electrons and the field reside is tangent to the ball of radius L in which the bound electrons and field reside. By averaging over rotations we may assume that the Coulomb potential seen by the n electrons is that of a point charge at the origin; since there is at least one of the outer balls that is a distance $2L$ from the origin, and since that ball contains at least one electron, we can safely say that the Coulomb interaction of the outer electrons with the

nuclei, i.e., the sum of the term (3.14) and the last term in (3.23), is less than $-(Z - N')/3L$. (The reason we wrote $Z - N'$ instead of $n(Z - N')$ is that we do not know the positions of the other $n - 1$ electrons; they could be very far away.)

To summarize the situation thus far, we have a negative Coulomb attraction of order CL^{-1} , where C is a fixed constant. We have various localization errors of order $L^{-2\gamma}$, R_0^{-2} and also $(R_0/L^\gamma)(L - 2R_0)^{-\gamma}(1 + |\log(\Lambda R_0)|)$. These latter terms can be made arbitrarily small compared to CL^{-1} if we choose $1 > \gamma > 3/4$ and

$$L^{2\gamma-1} \gg R_0 \gg L^{1/2}. \quad (3.24)$$

Finally, there are the terms (3.11) and (3.18 – 3.22) which involve expectation values of linear operators $a^\#$ in $\Phi(n)$ and in $\Psi(N')$. These are dangerous looking terms; on the face of it they appear to possibly be of order L^{-1} , but we can make them all effectively vanish!

To eliminate these terms we can make an anti-unitary transformation on $\Psi(N')$ (or else on $\Phi(n)$, but not on both) that will not alter the energy of each subunit or alter the Coulomb interaction. This anti-unitary is simply to replace $a^\#$ by $-a^\#$ and, simultaneously use complex conjugation to change $\Psi(N')$ to its complex conjugate $\bar{\Psi}(N')$. In addition we apply the unitary operator $\mathcal{W} = \prod_{i=1}^{N'} \sigma_i^{(2)}$, where σ^2 is the second Pauli matrix in the usual basis in which σ^2 has purely imaginary elements and σ^1 and σ^3 are real.

The effect of applying this anti-unitary is to replace (3.11) and (3.18 – 3.22) by their negatives, whereas all other energy terms remain unchanged. Note that the anti-unitary when applied to $\Psi(N')$ changes the sign of one of the factors in (3.18 – 3.22) only. It changes the sign of $(\Psi(N'), p_i \Psi(N'))$ because of the complex conjugation and it changes the sign of $(\Psi(N'), \sigma_i \Psi(N'))$ because of complex conjugation and $\mathcal{W}$. The terms $(\Psi(N'), A(x_i) \Psi(N'))$ and $(\Psi(N'), B(x_i) \Psi(N'))$ change their sign because of the change of sign of the $a^\#$'s. Thus, each of these terms can be negated, and one choice or the other will make *the sum* (but perhaps not each individual term) of (3.11) and (3.18 – 3.22) non-positive. $\square$

4 Localization Estimates for ‘Free’ Electrons and Photons

The main result of this section is Theorem 4.3 which shows how to construct a state in which the ‘free’ electrons and the field are localized. This was used

in Part 1 of the proof of Theorem 3.1. Part 2 of the proof of Theorem 3.1 also uses the part of Theorem 4.3 relating to the field localization.

The proofs in this section rely, in part, on the commutator estimates of Sects. 5 and 6. The Hamiltonian for the n free electrons is given in (2.3).

4.1 Localization of the Electrons, but not the Photons

LEMMA 4.1 (localization of electrons). *Fix a radius $R_0 > 0$. Then there exist (not necessarily disjoint) balls $B_1, \dots, B_n$ in $\mathbb{R}^3$, each of radius R_0 , and a normalized vector Ψ in the physical Hilbert space $\mathcal{H}(n)$ such that the electronic part of Ψ is supported in $\Omega(B_1, \dots, B_n)$ and with an energy*

$$(\Psi, H^0(n)\Psi) < E^0(n) + b n^2 R_0^{-2}, \tag{4.1}$$

where $b = 2\pi^2$ is twice the lowest eigenvalue of the Dirichlet Laplacian in a ball of radius 1.

Conjecture: The proof does not tell us the location of the n balls. If they happen to be distinct then we can replace n^2 by n in (4.1). We conjecture that the theorem can be generally improved in this way, i.e., $n^2 \rightarrow n$, with, perhaps, a different value for b .

PROOF: Let $\varepsilon = b n^2 R_0^{-2}/2$ and let Φ be a normalized approximate ground state with error at most $\varepsilon/2$, i.e., $\Phi \in \mathcal{H}$ and $(\Phi, H^0\Phi) < E^0(n) + \varepsilon/2$. Let B denote the ball of radius R_0 centered at the origin in $\mathbb{R}^3$ and let χ be a normalized, nonnegative, infinitely differentiable function with support in B . Define the function G of $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ by

$$G(X, Y) = \sum_{\pi \in S_n} \prod_{i=1}^n \chi(x_i - y_{\pi i}) , \tag{4.2}$$

where S_n is the symmetric group. Clearly G is a symmetric function of the X variables and of the Y variables and, therefore, $G(X, Y)\Phi$ is a valid vector in the physical Hilbert space for each choice of Y .

It is obvious that

$$P(X) := \int_{\mathbb{R}^{3n}} G(X, Y)^2 dY \tag{4.3}$$

is simply $n!$ times the permanent of the $n \times n$ hermitian, positive semidefinite matrix $M_{i,j} := \int_{\mathbb{R}^3} \chi(x_i - y)\chi(x_j - y)dy$. It is a general fact that such a

permanent is not less than the product of its diagonal elements; in fact it is not less than the product of the permanent of any principal $(n-m) \times (n-m)$ submatrix and the permanent of its $m \times m$ complement [11], so $P(X) \geq n!$. In particular, a fact that we shall use later is that for each i , $P(X) \geq n!M_{i,i}\widetilde{M}_{i,i} = n!\widetilde{M}_{i,i}$, where $\widetilde{M}_{i,i}$ is the permanent cofactor of $M_{i,i}$, i.e., it is the permanent of the matrix in which the i^{th} row and column is deleted from M . (In our case, the assertion is obvious since every matrix element $M_{i,j} > 0$.)

We define $W(X, Y) = G(X, Y)P(X)^{-1/2}$, and using this we define

$$\Psi_Y := W(X, Y)\Phi. \quad (4.4)$$

Our Ψ will be Ψ_Y (up to normalization) for a choice of Y to be determined shortly. Since $P(X) \geq n!$ the multiplier $W(X, Y)$ is C_c^∞ .

We proceed analogously to Theorem 3.1 of [7]. Consider

$$\mathcal{E}(Y) := (\Psi_Y, H^0(n)\Psi_Y) - [E^0(n) + \varepsilon + b n^2 R_0^{-2}] (\Psi_Y, \Psi_Y). \quad (4.5)$$

Our goal is to show that $\int \mathcal{E}(Y) dY < 0$ for a suitable choice of χ . This will prove that there is a set of Y 's of positive measure such that $\Psi_Y \neq 0$ and also $(\Psi_Y, H^0(n)\Psi_Y) / (\Psi_Y, \Psi_Y) \leq [E^0(n) + \varepsilon + b n^2 R_0^{-2}]$, which is what we wish to prove.

It is obvious, from (4.3) that $\int W(X, Y)^2 dY = 1$ and so

$$\int (\Psi_Y, \Psi_Y) dY = (\Phi, \Phi) = 1. \quad (4.6)$$

In a similar fashion one sees that

$$\int (\Psi_Y, [\alpha \sum_{i < j} |x_i - x_j|^{-1} + H_f] \Psi_Y) dY = (\Phi, [\alpha \sum_{i < j} |x_i - x_j|^{-1} + H_f] \Phi). \quad (4.7)$$

Next, we compute

$$\begin{aligned} \|(\nabla_{x_i} + iA(x_i))\Psi_Y\|^2 &= \|(\nabla_{x_i} W)\Phi\|^2 + \Re(\Phi, (\nabla_{x_i} W^2) \cdot (\nabla_{x_i} + iA(x_i))\Phi) \\ &\quad + \|W(\nabla_{x_i} + iA(x_i))\Phi\|^2. \end{aligned} \quad (4.8)$$

The middle term vanishes when we integrate over Y since $\int W(X, Y)^2 dY = 1$ and hence $\int \nabla_{x_i} W(X, Y)^2 dY = 0$. The last term gives us the required contribution of the kinetic energy to $\int (\Psi_Y, H^0(n)\Psi_Y) dY$ in (4.5), again using the fact that $\int W(X, Y)^2 dY = 1$. The first term is $(\Phi, F_i(X)\Phi)$, where

$$F_i(X) = \int \left| \nabla_{x_i} \left\{ G(X, Y)P(X)^{-1/2} \right\} \right|^2 dY. \quad (4.9)$$

Our proof is complete if we can show that $F_i(X) \leq 3\varepsilon/2n$, which we shall do next. We start with

$$\begin{aligned} \nabla_{x_i} \left\{ G(X, Y) P(X)^{-1/2} \right\} &= P(X)^{-1/2} \nabla_{x_i} G(X, Y) \\ &\quad - (1/2) G(X, Y) P(X)^{-3/2} \nabla_{x_i} P(X) . \end{aligned} \tag{4.10}$$

If we square this and integrate over Y we obtain (recalling that $\nabla_{x_i} P(X) = 2 \int G(X, Y) \nabla_{x_i} G(X, Y) dY$ and $\int G(X, Y)^2 dY = P(X)$)

$$F_i(X) = \frac{1}{P(X)} \int |\nabla_{x_i} G(X, Y)|^2 dY - \frac{1}{4P(X)^2} |\nabla_{x_i} P(X)|^2 . \tag{4.11}$$

We shall ignore the last term since it is negative.

In order to compute $\nabla_{x_i} G(X, Y)$ let us write

$$G(X, Y) = \sum_{j=1}^n \chi(x_i - y_j) \mu_j(X', Y') := \sum_{j=1}^n a_j(X, Y), \tag{4.12}$$

where

$$\mu_j(X', Y') = \sum_{\pi \in S_{n-1}} \prod_{\ell \neq i} \chi(x_\ell - y_{\pi \ell}) , \tag{4.13}$$

and where S_{n-1} denotes the set of bijections of $1, \dots, \hat{i}, \dots, n$ into $1, \dots, \hat{j}, \dots, n$. Then,

$$\begin{aligned} \int |\nabla_{x_i} G|^2 dY &= \sum_{j=1}^n \sum_{k=1}^n \int \nabla_{x_i} a_j \cdot \nabla_{x_i} a_k dY \leq n \sum_{j=1}^n \int |\nabla_{x_i} a_j|^2 dY \\ &= n \sum_{j=1}^n \int |(\nabla \chi)(x_i - y_j)|^2 dy_j \int \mu_j(X', Y')^2 dY' = n C_{i,i} \int |\nabla \chi(x)|^2 dx , \end{aligned} \tag{4.14}$$

where $Y' = (y_1, \dots, \hat{y}_j, \dots, y_n)$ and where $C_{i,i} = \int \mu(X', Y')^2 dY'$ equals $(n-1)!$ times $\widetilde{M}_{i,i}$, the cofactor of $M_{i,i}$ in the permanent of M . However, $P(X) \geq n! M_{i,i} \widetilde{M}_{i,i} = n! M_{i,i} [C_{i,i}/(n-1)!]$, as explained before. Therefore, $P^{-1} \int |\nabla_{x_i} G|^2 dY \leq n \int |\nabla \chi|^2$. The same inequality holds for any $i = 1, \dots, n$ which gives us a factor of n^2 altogether.

At this point we wish to choose χ to be the lowest Dirichlet eigenfunction of $-\Delta$ in the ball B . That function is not in $C_c^\infty(B)$, but it can be approximated by such a function so that the error in $n^2 \int |\nabla \chi|^2$ is less than $\varepsilon/2$. □

4.2 Localization of Photons

Our next task is to produce a state in which the photons are localized. A definition is needed first. We say that the electromagnetic field in a state Φ is **supported** in a closed subset $\Sigma \subset \mathbb{R}^3$ if each component a_λ of the field satisfies $\|a_\lambda(x)\Phi\| = 0$ for all $x \notin \Sigma$. To construct a localized state from any given state Φ we use the representation (2.15) of Fock space. We suppress the space and spin variables of the electrons for the moment. For a smooth cutoff function $0 \leq j(y) \leq 1$ define the localization operator $\mathcal{J}$ on Fock space in the following manner. $\mathcal{J}\Phi$ is still given by (2.15) but the vector $|i_1, m_1; \dots; i_n, m_n\rangle$ is changed to

$$\mathcal{J}|i_1, m_1; \dots; i_n, m_n\rangle = \frac{1}{\sqrt{m_1! \dots m_n!}} a^*(j f_{i_1})^{m_1} \dots a^*(j f_{i_n})^{m_n} |0\rangle, \quad (4.15)$$

Clearly, $\mathcal{J}$ is a linear, self-adjoint operator and

$$\|a_\lambda(y)\mathcal{J}\Phi\| = 0 \quad (4.16)$$

for all y that are outside the support of the function $j(y)$. Note that $\mathcal{J}$ is a contraction, i.e., $\|\mathcal{J}\Phi\| \leq \|\Phi\|$ for all Φ .

To work effectively with the operator $\mathcal{J}$ the following commutation relations will be useful later. Since the electron variables are not relevant for the calculation, we suppress them here.

LEMMA 4.2 (Commutation relations for $\mathcal{J}$). *For any f in the single photon Hilbert space $L^2(\mathbb{R}^3; \mathbb{C}^2)$ we have (with $a_\lambda(f) = \int a_\lambda(y) \overline{f(y)} dy$, as before) that*

$$a_\lambda(f)\mathcal{J} = \mathcal{J}a_\lambda(jf), \quad \mathcal{J}a_\lambda^*(f) = a_\lambda^*(jf)\mathcal{J} \quad (4.17)$$

$$[a_\lambda(f), \mathcal{J}] = \mathcal{J}a_\lambda((j-1)f), \quad [a_\lambda^*(f), \mathcal{J}] = -a_\lambda^*((j-1)f)\mathcal{J}. \quad (4.18)$$

PROOF: For any state Ψ in the Fock space we have that

$$\begin{aligned} [a_\lambda(f)\mathcal{J}\Psi]_n(y_1, \lambda_1; \dots; y_n, \lambda_n) \\ = \sqrt{n+1} \prod_{k=1}^n j(y_k) \int \overline{f(y)} j(y) [\Psi]_{n+1}(y, \lambda; y_1, \lambda_1; \dots; y_n, \lambda_n) dy \\ = [\mathcal{J}a_\lambda(jf)\Psi]_n(y_1, \lambda_1; \dots; y_n, \lambda_n). \end{aligned} \quad (4.19)$$

All the relations follow immediately from this. $\square$

4.3 Localization of the Photons and the Electrons Together

One would like to think that most of the photons ought to be localized near the electrons. This will be the case provided one replaces the state Ψ of Lemma 4.1 by the ground state for the Hamiltonian $H^0(n)$ restricted to the states that vanish outside the set Ω . Moreover, this state will have an energy close to the energy $E^0(n)$. The following theorem makes this precise.

THEOREM 4.3 (Localized photons and free electrons). *Fix radii $R_0 > 0$ and $L > 2R_0$. Then there exist (not necessarily disjoint) balls $B_1, \dots, B_n$ in $\mathbb{R}^3$, each of radius R_0 and a normalized vector $\Phi(n)$ in the physical Hilbert space $\mathcal{H}(n)$ such that the electronic part of $\Phi(n)$ is supported in $\Omega(B_1, \dots, B_n)$ and the electromagnetic field is supported in $\Sigma = \cup_{i=1}^n P_i$ where P_i is a ball concentric with B_i but with radius L .*

The energy of $\Phi(n)$ satisfies

$$(\Phi(n), H^0(n)\Phi(n)) < E^0(n) + b \frac{n^2}{R_0^2} + c \frac{n}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L} \right) (1 + |\log(\Lambda R_0)|), \quad (4.20)$$

for any $\gamma < 1$ and where $b = 2\pi^2$ is twice the lowest eigenvalue of the Dirichlet Laplacian in a ball of radius 1. The constant c depends only on γ and is independent of R_0 and L .

PROOF: We start with the wave function Ψ given by Lemma 4.1. This fixes the balls $B_1, \dots, B_n$, and hence the symmetrized product $\Omega(B_1, \dots, B_n)$. The next step is to redefine the Hamiltonian $H^0(n)$ by restricting the Hilbert space to the balls, i.e., we replace the space $\wedge L^2(\mathbb{R}^3; \mathbb{C}^2)$ by the subspace of $L^2(\Omega; \otimes_1^n \mathbb{C}^2)$ consisting of functions that are antisymmetric under the exchange of particle labels. (This makes sense because Ω is symmetric under exchange of particle coordinates.) The Laplacian is replaced by the Dirichlet Laplacian.

A physical way to say this is that we add an infinite potential outside Ω . This is not a sum of single particle potentials, but that is immaterial. The point is that by the methods of [7] there is a bound state, i.e., there is a state $\Phi_D(n)$ with lowest energy $E_D^0(n)$ (the letter D stands for ‘Dirichlet’) that satisfies Schrödinger’s equation. (In fact, the methods of [7] are not needed to establish the existence of a ground state in this case since all finite energy states are evidently localized; this was noted earlier in [5], [1] and [8]. However, [7] is needed for the photon localization in the next step.)

This ground state will obviously have a lower energy than the Ψ given by

Lemma 4.1 since that Ψ automatically satisfies the Dirichlet boundary conditions, that is $(\Phi_D(n), H^0(n)\Phi_D(n)) \leq (E^0(n) + bn^2R_0^{-2})(\Phi_D(n), \Phi_D(n))$.

Next, we localize the photons in the set Σ . A standard IMS localization yields two smooth functions $j_1(y), j_2(y)$ with

$$j_1(y)^2 + j_2(y)^2 = 1 \quad (4.21)$$

with support of $j_1(y)$ in Σ . We also require that $j_1(y)$ is identically equal to 1 on the set $\cup_{i=1}^n Q_i$ where, for each i , Q_i is a ball of radius $L/2$, concentric with P_i . Moreover we can assume that $|\nabla j_i(y)| \leq C/L$ for some constant C and $i = 1, 2$. We define $\mathcal{J}\Phi_D(n)$ by using j_1 in (4.15) and, with the help of (4.16), we use the localized state $\mathcal{J}\Phi_D(n)$, appropriately normalized, as a trial function. This function will be the required function $\Phi(n)$ of our theorem.

The energy of the state $\mathcal{J}\Phi_D(n)$ can be compared with the energy of $\Phi_D(n)$ by using the commutator formula

$$(\mathcal{J}\Phi_D(n), (H^0(n) - E_D^0(n))\mathcal{J}\Phi_D(n)) = (\mathcal{J}\Phi_D(n), [H^0(n), \mathcal{J}]\Phi_D(n)) \quad (4.22)$$

An important point about having a *ground state* $\Phi_D(n)$ is that one can derive infrared bounds for this state (see Sects. 5 and 6). All that is needed is that Φ_D is a ground state, i.e., it satisfies the Schrödinger equation in order to apply the ‘pull through formula’.

Lemma 5.2 shows that the norm $\|\mathcal{J}\Phi_D\|$ is close to one and Lemma 5.5 shows that the right side of (4.22) is bounded as

$$\begin{aligned} & (\mathcal{J}\Phi_D(n), [H^0(n), \mathcal{J}]\Phi_D(n)) \\ & \leq \frac{C}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L^\gamma} \right) (1 + |\log(\Lambda R_0)|) (\mathcal{J}\Phi_D(n), \mathcal{J}\Phi_D(n)) \quad , \quad (4.23) \end{aligned}$$

for any $\gamma < 1$, where the constant C depends on γ but not on R_0 and L . This shows that $\Phi(n) = \mathcal{J}\Phi_D(n)/\|\mathcal{J}\Phi_D(n)\|$ satisfies (4.20). $\square$

5 Commutator and Related Estimates

In this section we prove various results stated in the previous sections, particularly Lemma 5.5 which is used in the proof of Theorem 4.3. We also prove Lemma 5.6, which is not a commutator estimate; it is simpler. It is

needed to bound the terms (3.16) and (3.17), as we stated just after eq. (3.23).

We will deal mainly with the Dirichlet ground state associated with electrons localized in the set $\Omega(B_1, \dots B_n)$. The bounds for the ground state describing electrons exponentially localized near the nuclei is easier and follows in the same fashion.

In this section and the next we denote the Dirichlet ground state $\Phi_D(n)$ simply by Φ in order to simplify the notation.

Recall the definitions of j_1, j_2 in (4.21) and of the operator $\mathcal{J}$ in (4.15) which is defined by substituting j_1 for j . An important operator in our analysis is the outer photon number, given by

$$\mathcal{N}_{\text{out}} = \sum_{\lambda=1}^2 \int_{j_2(y)>0} a_{\lambda}^*(y) a_{\lambda}(y) dy \tag{5.1}$$

or, in terms of matrix elements,

$$(\Phi, \mathcal{N}_{\text{out}} \Phi) := \sum_{\lambda=1}^2 \int_{j_2(y)>0} \|a_{\lambda}(y) \Phi\|^2 dy \tag{5.2}$$

We start with the following bound, which is a consequence of the infrared bounds proved in Sect. 6. It is used in the proofs of Lemmas 5.2, 5.3 and 5.4.

LEMMA 5.1 (Photon number is small far away from the electrons). *For the Dirichlet ground state Φ of the free electrons localized in $\Omega(B_1, \dots B_n)$ we have the bound (with C independent of R_0 and L)*

$$(\Phi, \mathcal{N} \Phi) \leq C(1 + |\log(\Lambda R_0)|) \tag{5.3}$$

and for all $\gamma < 1$, the bound

$$(\Phi, \mathcal{N}_{\text{out}} \Phi) \leq C \left(\frac{R_0}{L^{\gamma}} \right)^2 \|\Phi\|^2 \tag{5.4}$$

where the constant C depends on n, Λ, γ but not on R_0 and L . Likewise, for the ground state Ψ of the bound system given by the Hamiltonian $H^V(N')$ we have that (with C independent of R_0 and L)

$$(\Psi, \mathcal{N} \Psi) \leq C \tag{5.5}$$

and for all $\gamma < 1$ and $L > 2R_0$ that

$$(\Psi, \mathcal{N}_{\text{out}} \Psi) \leq C \left(\frac{1}{L^{2\gamma}} \right) \|\Psi\|^2 \tag{5.6}$$

where the constant C depends only on γ .

Our goal is to prove inequality (4.23), which is Lemma 5.5 of this section. To do so, we shall need the following three lemmas, in which $\mathcal{J}$ is defined with j_1 . Recall that $0 \leq j_1(y) \leq 1$ and $j_1(y) = 1$ for $y \in \bigcup_{i=1}^n Q_i$ (see the proof of Theorem 4.3).

LEMMA 5.2. *For the normalized ground state Φ , we have for all $0 < \gamma < 1$ that*

$$1 - \|\mathcal{J}\Phi\|^2 \leq C \left(\frac{R_0}{L^\gamma} \right)^2 \quad (5.7)$$

where C is a constant that depends on γ, n, Λ but not on R_0 and L . Moreover, for an arbitrary state Ψ ,

$$(\mathcal{J}\Psi, \mathcal{N}_{\text{out}} \mathcal{J}\Psi) \leq (\Psi, \mathcal{N}_{\text{out}} \Psi) . \quad (5.8)$$

PROOF: Formula (5.8) is immediate from

$$(\mathcal{J}\Psi, \mathcal{N}_{\text{out}} \mathcal{J}\Psi) = \sum_{n=1}^{\infty} n \left\| \prod_{l=1}^n j(y_l) \prod_{l=1}^n \chi_{j_2(y_l) > 0} [\Psi]_n \right\|^2 . \quad (5.9)$$

Next, note that

$$1 - \prod_{k=1}^n j_1^2(y_k) = \sum_{l=1}^n \prod_{k=1}^{l-1} j_1^2(y_k) j_2^2(y_l) \quad (5.10)$$

(by definition the empty product equals 1). This is proved by inserting $j_2^2(y_n) = 1 - j_1^2(y_n)$ on the left side of (5.10) and then repeating the process inductively. In particular, we have that

$$1 - \prod_{k=1}^n j_1^2(y_k) \leq \sum_{l=1}^n j_2^2(y_l) , \quad (5.11)$$

from which we obtain

$$1 - \|\mathcal{J}\Phi\|^2 \leq (\Phi, \mathcal{N}_{\text{out}} \Phi) \leq C \left(\frac{R_0}{L^\gamma} \right)^2 \quad (5.12)$$

by Lemma 5.1. □

LEMMA 5.3. *For the ground state Φ and for every $L > 2R_0$ we have that*

$$\begin{aligned} & \left| \left(\mathcal{J}\Phi, \sum_{i=1}^n [(p_i + A(x_i))^2, \mathcal{J}] \Phi \right) \right| \\ & \leq C \frac{1}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L^\gamma} \right) (1 + |\log(\Lambda R_0)|) \|\mathcal{J}\Phi\|^2 , \end{aligned} \quad (5.13)$$

for all $\gamma < 1$. C is a constant that depends on γ, n, Λ but not on R_0 and L . (Note that it makes no difference whether we use $(p_i + A(x_i))^2$ or use its normal ordering since the commutator of a_i and a_i^* is proportional to the identity operator, which commutes with $\mathcal{J}$.)

PROOF: We first calculate the commutator of $(p + A(x))^2$ with $\mathcal{J}$.

$$\begin{aligned} \sum_{i=1}^n [(p_i + A(x_i))^2, \mathcal{J}] = \\ \sum_{i=1}^n (2p_i \cdot ([a_i, \mathcal{J}] + [a_i^*, \mathcal{J}]) + [a_i a_i, \mathcal{J}] + [a_i^* a_i^*, \mathcal{J}] + 2[a_i^* a_i, \mathcal{J}]) \quad , \quad (5.14) \end{aligned}$$

where we abbreviated $a_\lambda(h_\lambda^j(x_i - \cdot))$ by a_i and likewise for a_i^* . (Note that the index $j \in \{1, 2, 3\}$ of the coupling function is unimportant and will be suppressed from now on.)

Step 1. The term $\sum_{i=1}^n (\mathcal{J}\Phi, 2p_i \cdot [a_i, \mathcal{J}]\Phi)$ is bounded, by Schwarz's inequality, by

$$2 \left(\sum_{i=1}^n \|p_i \mathcal{J}\Phi\|^2 \right)^{1/2} \left(\sum_{i=1}^n \|[a_i, \mathcal{J}]\Phi\|^2 \right)^{1/2} . \quad (5.15)$$

The first factor can be estimated simply in terms of the energy while the second factor will deliver the necessary decay in L . Using Lemma 4.2 the problem is reduced to estimating, for each fixed $X = (x_1, \dots, x_n) \in \Omega$ and each i (with $\|\cdot\|$ denoting the norm in Fock space only)

$$\begin{aligned} & \|\mathcal{J}a(h(x_i - \cdot)(j_1(\cdot) - 1))\Phi\| \\ & \leq \|a(h(x_i - \cdot)(j_1(\cdot) - 1))\Phi\| \leq \int (1 - j_1(y)) \|h(x_i - y)a(y)\Phi\| dy \\ & \leq \int (1 - j_1(y)) |h(x_i - y)| \|a(y)\Phi\| dy \\ & \leq \left(\int (1 - j_1(y)) |h(x_i - y)|^2 dy \right)^{1/2} \left(\int (1 - j_1(y)) \|a(y)\Phi\|^2 dy \right)^{1/2} . \end{aligned} \quad (5.16)$$

The first factor in (5.16) can be bounded, using Lemma B.1, by

$$\left(\int (1 - j_1(y)) \frac{1}{(L - 2R_0)^{2\gamma}} |x_i - y|^{2\gamma} |h(x_i - y)|^2 dy \right)^{1/2} \leq \frac{C}{(L - 2R_0)^\gamma} \quad (5.17)$$

since, whenever $1 - j_1(y) \neq 0$, the distance between x_i and y is at least $d = (L/2) - R_0$, by construction. If we now square (5.16) and integrate over X we get the desired decay estimate for (5.15).

For the second factor in (5.16) we note that $1 - j_1(y) \leq j_2(y)^2$. This, together with Lemma 5.1, yields

$$\left(\sum_{i=1}^n \|\mathcal{J}a(h(x_i - \cdot)(j_1(\cdot) - 1))\Phi\|^2 \right)^{1/2} \leq \frac{C}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L^\gamma} \right) \quad (5.18)$$

for all $\gamma < 1$.

Next, $(\mathcal{J}\Phi, 2 \sum_{i=1}^n p_i \cdot [a_i^*, \mathcal{J}]\Phi) = -2 \sum_{i=1}^n ([a_i, \mathcal{J}]\mathcal{J}\Phi, p_i\Phi)$ and this can be estimated in the same fashion as before except that the estimate is in terms of

$$(\mathcal{J}\Phi, \mathcal{N}_{\text{out}}\mathcal{J}\Phi)$$

instead of $(\Phi, \mathcal{N}_{\text{out}}\Phi)$. On account of Lemma 5.2, this is bounded by

$$(\Phi, \mathcal{N}_{\text{out}}\Phi) .$$

Hence, we obtain the same kind of bound as in (5.18), i.e., for all $\gamma < 1$,

$$\left| \sum_{i=1}^n (\mathcal{J}\Phi, 2p_i \cdot [a_i^*, \mathcal{J}]\Phi) \right| \leq \frac{C}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L^\gamma} \right) .$$

Step 2. Returning to (5.14) we concentrate on the term $[a_i a_i, \mathcal{J}]$ which can be written as $a_i[a_i, \mathcal{J}] + [a_i, \mathcal{J}]a_i$. Using Schwarz's inequality

$$(\mathcal{J}\Phi, a_i[a_i, \mathcal{J}]\Phi) = (a_i^* \mathcal{J}\Phi, [a_i, \mathcal{J}]\Phi) \leq \|a_i^* \mathcal{J}\Phi\| \|[a_i, \mathcal{J}]\Phi\| . \quad (5.19)$$

The second factor is treated in precisely the same fashion as in Step 1. The first factor cannot be estimated directly in terms of the energy, since the function $\mathcal{J}\Phi$ is not an eigenfunction. This will be dealt with below where we estimate the term $\|a^* \mathcal{J}^2 \Phi\|$.

The term

$$(\mathcal{J}\Phi, [a_i, \mathcal{J}]a_i\Phi) , \quad (5.20)$$

can be written, using Lemma 4.2, as

$$\begin{aligned} & (\mathcal{J}\Phi, \mathcal{J}a(h(x_i - \cdot)(j_1(\cdot) - 1))a(h(x_i - \cdot))\Phi) \\ &= (\mathcal{J}^2\Phi, a(h(x_i - \cdot))a(h(x_i - \cdot)(j_1(\cdot) - 1))\Phi) \end{aligned} \quad (5.21)$$

which, again using Schwarz's inequality, can be bounded by

$$\|a^*(h(x_i - \cdot))\mathcal{J}^2\Phi\| \|a(h(x_i - \cdot)(j_1(\cdot) - 1))\Phi\| . \tag{5.22}$$

As before, the first factor cannot be estimated in terms of the energy, since $\mathcal{J}^2\Phi$ is not an eigenstate. Note, however, that

$$\|a^*\mathcal{J}^2\Phi\|^2 = \|a\mathcal{J}^2\Phi\|^2 + \|h\|^2\|\mathcal{J}^2\Phi\|^2 . \tag{5.23}$$

and the first term on the right side can be estimated by

$$\|h\|^2 (\mathcal{J}^2\Phi, \mathcal{N}\mathcal{J}^2\Phi) \leq \|h\|^2 (\Phi, \mathcal{N}\Phi) . \tag{5.24}$$

This follows from the formula

$$\mathcal{N} = \sum_{j=1}^{\infty} a^*(f_j)a(f_j) , \tag{5.25}$$

which is valid for any orthonormal basis $\{f_j\}$, in which we pick $f_1(y) = h(x - y)/\|h(x - \cdot)\|$, and from (5.9). (Here h is an abbreviation for the coupling functions.) Thus, using Lemma 5.1,

$$\left| \sum_{i=1}^n (\mathcal{J}\Phi, [a_i a_i, \mathcal{J}]\Phi) \right| \leq C \frac{1}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L^\gamma} \right) (1 + |\log(\Lambda R_0)|) . \tag{5.26}$$

Step 3. By taking adjoints the third term in (2), $\sum_{i=1}^n [a_i^* a_i^*, \mathcal{J}]$ leads to the expression

$$- \sum_{i=1}^n ([a_i a_i, \mathcal{J}]\mathcal{J}\Phi, \Phi) \tag{5.27}$$

and can be dealt with in the same fashion as in Step 2. It remains to analyze $\sum_{i=1}^n [a_i^* a_i, \mathcal{J}] = \sum_{i=1}^n a_i^* [a_i, \mathcal{J}] + [a_i^*, \mathcal{J}] a_i$. The first term is estimated using

$$(a_i \mathcal{J}\Phi, [a_i, \mathcal{J}]\Phi) \leq \|a_i \mathcal{J}\Phi\| \|[a_i, \mathcal{J}]\Phi\| , \tag{5.28}$$

while the second one can be written as

$$- ([a_i, \mathcal{J}]\mathcal{J}\Phi, a_i \Phi) \tag{5.29}$$

which, once more by Schwarz's inequality, can be bounded by

$$\|[a_i, \mathcal{J}]\mathcal{J}\Phi\| \|a_i \Phi\| . \tag{5.30}$$

Both these terms have been estimated previously. □

We come now to the third lemma needed for the proof of Lemma 5.5. This lemma concerns only the real part of a commutator expectation value (5.31), but this is all we need for Lemma 5.5. The reason is that the total commutator in Lemma 5.5 is manifestly real, since $\Phi(n)$ is an eigenstate of $H^0(n)$ and $\mathcal{J}$ is selfadjoint. On the other hand, the piece of the commutator considered in Lemma 5.3 is also manifestly real and the only other part of $H^0(n)$ to be considered is the potential energy terms, which commute with $\mathcal{J}$. Therefore, the commutator expectation value in (5.31) is, in fact, real. The proof of Lemma 5.4 is greatly simplified, however, by being able to ignore the (non-existent) imaginary part.

LEMMA 5.4 (Commutator of $\mathcal{J}$ with the field energy). *The ground state Φ satisfies the bound*

$$\Re(\mathcal{J}\Phi, [H_f, \mathcal{J}]\Phi) \leq \frac{C}{L} \left(\frac{R_0}{L^\gamma} \right) (1 + |\log(\Lambda R_0)|)^{1/2} \quad (5.31)$$

for any $\gamma < 1$. C is a constant that depends on γ, n, Λ but not on R_0 and L .

PROOF: It is convenient to write the field energy of a state Φ in the form

$$\begin{aligned} (\Phi, H_f \Phi) &= (2\pi)^3 \sum_{\lambda} \left(a_{\lambda}(\cdot) \Phi, \sqrt{-\Delta} a_{\lambda}(\cdot) \Phi \right) = \\ &= 4\pi \sum_{\lambda} \int \frac{\|a_{\lambda}(x)\Phi - a_{\lambda}(y)\Phi\|^2}{|x - y|^4} dx dy, \end{aligned} \quad (5.32)$$

(see [10, Eq. 7.12(4)]). Next, we note that the commutator expression (5.31) is given by

$$\begin{aligned} (\mathcal{J}\Phi, H_f \mathcal{J}\Phi) - \Re(\mathcal{J}\Phi, \mathcal{J}H_f \Phi) &= 4\pi \sum_{\lambda} \int \frac{\|a_{\lambda}(x)\mathcal{J}\Phi - a_{\lambda}(y)\mathcal{J}\Phi\|^2}{|x - y|^4} dx dy \\ &- 4\pi \sum_{\lambda} \Re \int \frac{(a_{\lambda}(x)\mathcal{J}^2\Phi - a_{\lambda}(y)\mathcal{J}^2\Phi, a_{\lambda}(x)\Phi - a_{\lambda}(y)\Phi)}{|x - y|^4} dx dy. \end{aligned} \quad (5.33)$$

First, we investigate the numerator of the sum of the two integrands, which is

$$\begin{aligned} (a_{\lambda}(x)\mathcal{J}\Phi, a_{\lambda}(x)\mathcal{J}\Phi) - \Re(a_{\lambda}(x)\mathcal{J}^2\Phi, a_{\lambda}(x)\Phi) \\ - \Re(a_{\lambda}(x)\mathcal{J}\Phi, a_{\lambda}(y)\mathcal{J}\Phi) + \Re(a_{\lambda}(x)\mathcal{J}^2\Phi, a_{\lambda}(y)\Phi) \end{aligned} \quad (5.34)$$

plus the same thing with y and x exchanged. For brevity's sake we have omitted the sums over λ here and in the following. We note that the x - x terms (and likewise the y - y terms) cancel. This follows from

$$\begin{aligned} & (a_\lambda(x)\mathcal{J}\Phi, a_\lambda(x)\mathcal{J}\Phi) - \Re(a_\lambda(x)\mathcal{J}^2\Phi, a_\lambda(x)\Phi) = \\ & (a_\lambda(x)\mathcal{J}\Phi, a_\lambda(x)\mathcal{J}\Phi) - \Re(\mathcal{J}a_\lambda(x)\mathcal{J}\Phi, a_\lambda(x)\Phi) - \Re([a_\lambda(x), \mathcal{J}]\mathcal{J}\Phi, a_\lambda(x)\Phi) \\ & = \Re(a_\lambda(x)\mathcal{J}\Phi, [a_\lambda(x), \mathcal{J}]\Phi) - \Re([a_\lambda(x), \mathcal{J}]\mathcal{J}\Phi, a_\lambda(x)\Phi) , \end{aligned} \quad (5.35)$$

which, together with Lemma 4.2, yields

$$\begin{aligned} & (j_1(x) - 1)\Re(a_\lambda(x)\mathcal{J}\Phi, \mathcal{J}a_\lambda(x)\Phi) - (j_1(x) - 1)\Re(\mathcal{J}a_\lambda(x)\mathcal{J}\Phi, a_\lambda(x)\Phi) \\ & = (j_1(x) - 1)[\Re(a_\lambda(x)\mathcal{J}\Phi, \mathcal{J}a_\lambda(x)\Phi) - \Re(a_\lambda(x)\mathcal{J}\Phi, \mathcal{J}a_\lambda(x)\Phi)] = 0 . \end{aligned} \quad (5.36)$$

Now we deal with last two terms in (5.34).

$$\begin{aligned} & \Re(a_\lambda(x)\mathcal{J}^2\Phi, a_\lambda(y)\Phi) - \Re(a_\lambda(x)\mathcal{J}\Phi, a_\lambda(y)\mathcal{J}\Phi) \\ & = \Re([a_\lambda(x), \mathcal{J}]\mathcal{J}\Phi, a_\lambda(y)\Phi) + \Re(a_\lambda(x)\mathcal{J}\Phi, \mathcal{J}a_\lambda(y)\Phi) \\ & \quad - \Re(a_\lambda(x)\mathcal{J}\Phi, a_\lambda(y)\mathcal{J}\Phi) \\ & = \Re([a_\lambda(x), \mathcal{J}]\mathcal{J}\Phi, a_\lambda(y)\Phi) + \Re(a_\lambda(x)\mathcal{J}\Phi, [\mathcal{J}, a_\lambda(y)]\Phi) . \end{aligned} \quad (5.37)$$

Again, by Lemma 4.2, this equals

$$(j_1(x) - 1)\Re(a_\lambda(x)\mathcal{J}\Phi, \mathcal{J}a_\lambda(y)\Phi) - (j_1(y) - 1)\Re(a_\lambda(x)\mathcal{J}\Phi, \mathcal{J}a_\lambda(y)\Phi) . \quad (5.38)$$

Commuting the $\mathcal{J}$ once more with the a 's leads to

$$\begin{aligned} & (j_1(x) - 1)\Re([a_\lambda(x), \mathcal{J}]\Phi, \mathcal{J}a_\lambda(y)\Phi) - (j_1(y) - 1)\Re([a_\lambda(x), \mathcal{J}]\Phi, \mathcal{J}a_\lambda(y)\Phi) \\ & \quad + \{j_1(x) - j_1(y)\}\Re(\mathcal{J}a_\lambda(x)\Phi, \mathcal{J}a_\lambda(y)\Phi) . \end{aligned} \quad (5.39)$$

We do not have to worry about the last term, since it is the product of a symmetric and an antisymmetric term in the variables x and y , and thus its x - y integral with $|x - y|^{-4}$ in (7) vanishes. The other term, using Lemma 4.2, is of the form

$$\{(j_1(x) - 1)^2 - (j_1(y) - 1)(j_1(x) - 1)\}\Re(\mathcal{J}a_\lambda(x)\Phi, \mathcal{J}a_\lambda(y)\Phi) . \quad (5.40)$$

Taking into account the terms with x and y exchanged we find that (5.33) equals

$$4\pi \sum_\lambda \int \frac{(j_1(x) - j_1(y))^2 \Re(\mathcal{J}a_\lambda(x)\Phi, \mathcal{J}a_\lambda(y)\Phi)}{|x - y|^4} dx dy . \quad (5.41)$$

Now we use Lemma 5.1 to get an estimate on the size of (5.41). Recall that the function $j_1(x)$, which defines $\mathcal{J}$ and which is defined in (4.21), is identically equal to 1 on $\cup_{i=1}^n Q_i$, where Q_i is the ball of radius $L/2$ equi-centered with the ball B_i . Moreover $j_1(x) = 0$ whenever the distance of x to the center of every B_i exceeds L . Write $1 = \chi_1 + \chi_2 + \chi_3$ where χ_1 is the characteristic function of $\cup_{i=1}^n Q_i$ and χ_2 is the characteristic function of the shell between $\cup_{i=1}^n Q_i$ and $\Sigma = \cup_{i=1}^n P_i$. Finally, χ_3 is the characteristic function of the outside region (on which $j_2(x) = 1$ and $j_1(x) = 0$). We note, for later use, that $\int \chi_1 \leq CnL^3$ and $\int \chi_2 \leq CnL^3$, where C is a universal constant.

Next, we analyze each of the terms

$$T_{i,j} = 4\pi \int \frac{(j_1(x) - j_1(y))^2 \chi_i(x) \chi_j(y) \Re(\mathcal{J}a_\lambda(x)\Phi, \mathcal{J}a_\lambda(y)\Phi)}{|x - y|^4} dx dy. \quad (5.42)$$

Clearly, $T_{1,1} = T_{3,3} = 0$. To bound the other $T_{i,j}$'s we recall that

$$(j_1(x) - j_1(y))^2 \leq \frac{C}{L^2} |x - y|^2. \quad (5.43)$$

With this and Schwarz's inequality, $T_{i,j}$ is bounded by

$$\begin{aligned} T_{i,j} &\leq \frac{C}{L^2} \int \frac{\chi_i(x) \|\mathcal{J}a_\lambda(x)\Phi\| \chi_j(y) \|\mathcal{J}a_\lambda(y)\Phi\|}{|x - y|^2} dx dy \\ &\leq \frac{C}{L^2} \int \frac{\chi_i(x) \|a_\lambda(x)\Phi\| \chi_j(y) \|a_\lambda(y)\Phi\|}{|x - y|^2} dx dy. \end{aligned} \quad (5.44)$$

Denote $\|a_\lambda(x)\Phi\|$ by $f(x)$. Consider the terms $i = 3$ and $j = 1, 2$. Using the Hardy-Littlewood-Sobolev inequality (see [10, Theorem 4.3] with $1/2 + 5/6 + 2/3 = 2$), we get the bound

$$T_{3,j} \leq \frac{C}{L^2} \|\chi_3 f\|_2 \|\chi_j f\|_{6/5}, \quad j = 1, 2. \quad (5.45)$$

By Hölder's inequality

$$\|\chi_j f\|_{6/5} \leq \|\chi_j\|_3 \|f\|_2 \leq Cn^{1/3} L \|f\|_2, \quad (5.46)$$

and hence, for $j = 1, 2$,

$$T_{3,j} \leq \frac{Cn^{1/3}}{L} (\Phi, \mathcal{N}\Phi)^{1/2} (\Phi, \mathcal{N}_{\text{out}}\Phi)^{1/2}. \quad (5.47)$$

Note that $\|\chi_3 f\|_2$ is proportional to $(\Phi, \mathcal{N}_{\text{out}}\Phi)^{1/2}$ while $\|f\|_2$ is proportional to $(\Phi, \mathcal{N}\Phi)^{1/2}$.

The next term to consider is $i = 2$ and $j = 1, 2, 3$. Again, H-L-S leads to the bound

$$T_{2,j} \leq \frac{C}{L^2} \|\chi_j f\|_2 \|\chi_2 f\|_{6/5}, \quad j = 1, 2, 3. \quad (5.48)$$

Applying Hölder's inequality yields

$$\|\chi_2 f\|_{6/5} \leq \|\chi_2\|_3 \|\chi_2 f\|_2 \leq Cn^{1/3} L \|\chi_2 f\|_2, \quad (5.49)$$

and hence the term with $i = 2$ and $j = 1, 2, 3$ is also bounded above by

$$\frac{Cn^{1/3}}{L} (\Phi, \mathcal{N}\Phi)^{1/2} (\Phi, \mathcal{N}_{\text{out}}\Phi)^{1/2}, \quad (5.50)$$

where we used $(\Phi, \mathcal{N}_{\text{out}}\Phi)^{1/2} \geq \|\chi_2 f\|_2$. It is the term $(\Phi, \mathcal{N}\Phi)$ which yields the logarithmic term in formula (5.31).

Finally, the term $i = 1, j = 2$ is the same as $i = 2, j = 1$, and the term $i = 1, j = 3$ is the same as $i = 3, j = 1$, both of which have been already treated.

Collecting the estimates we have shown that

$$\Re(\mathcal{J}\Phi, [H_f, \mathcal{J}]\Phi) \leq \frac{Cn^{1/3}}{L} (\Phi, \mathcal{N}\Phi)^{1/2} (\Phi, \mathcal{N}_{\text{out}}\Phi)^{1/2}$$

which, by Lemma 5.1, proves the lemma. $\square$

We are now ready to prove the estimate stated at the end of the proof of Theorem 4.3, i.e.,

LEMMA 5.5. *For all $L > 2R_0$ we have the estimate*

$$\frac{(\mathcal{J}\Phi(n), [H^0(n), \mathcal{J}]\Phi(n))}{(\mathcal{J}\Phi(n), \mathcal{J}\Phi(n))} \leq \frac{C}{(L - 2R_0)^\gamma} \left(\frac{R_0}{L}\right)^\gamma (1 + |\log(\Lambda R_0)|) \quad (5.51)$$

for all $\gamma < 1$. The constant C depends on n, Λ, γ but not on R_0 and L .

PROOF: The lemma is a direct consequence of Lemmas 5.2, 5.3 and 5.4. $\square$

LEMMA 5.6 (Bound on the error terms). *For each fixed $x \in \mathbb{R}^3$,*

$$(\mathcal{J}\Phi, a(h(x - \cdot))^2 \mathcal{J}\Phi) \leq C_\gamma \frac{1}{D^{2\gamma}} \max\{(\Phi, \mathcal{N}\Phi), 1\}, \quad (5.52)$$

for any $\gamma < 1$, where D is the distance of x to the support of j_1 . The same estimate holds for $a^*(h(x - \cdot))^2$ and for $a^*(h(x - \cdot))a(h(x - \cdot))$ in place of $a(h(x - \cdot))^2$.

PROOF: Using Lemma 4.2,

$$(\mathcal{J}\Phi, a(h(x - \cdot))^2 \mathcal{J}\Phi) = (\mathcal{J}\Phi, \mathcal{J}a(h(x - \cdot)j_1(\cdot))^2 \Phi) = (a^*(h(x - \cdot)j_1(\cdot))\mathcal{J}^2\Phi, a(h(x - \cdot)j_1(\cdot))\Phi). \quad (5.53)$$

By Schwarz's inequality this is bounded above by

$$\|(a^*(h(x - \cdot)j_1(\cdot))\mathcal{J}^2\Phi)\| \|a(h(x - \cdot)j_1(\cdot))\Phi\|. \quad (5.54)$$

Similar to the proof of Lemma 5.3, the second factor in (5.54) can be bounded as follows.

$$\begin{aligned} \|a(h(x - \cdot)j_1(\cdot))\Phi\| &\leq \int j_1(y) \|h(x - y)a(y)\Phi\| dy \\ &\leq \int j_1(y) |h(x - y)| \|a(y)\Phi\| dy \\ &\leq \left(\int j_1(y)^2 |h(x - y)|^2 dy \right)^{1/2} \left(\int \|a(y)\Phi\|^2 dy \right)^{1/2}. \end{aligned} \quad (5.55)$$

The second factor here is $(\Phi, \mathcal{N}\Phi)^{1/2}$ while the first factor can be estimated, using the fact that the support of j_1 and the point x are a distance D apart, as

$$\frac{1}{D^\gamma} \left(\int j_1(y)^2 |x - y|^{2\gamma} |h(x - y)|^2 dy \right)^{1/2} \leq \frac{C}{D^\gamma} \quad (5.56)$$

by Lemma B.1.

The first factor in (5.54) can be bounded as follows.

$$\begin{aligned} \|(a^*(h(x - \cdot)j_1(\cdot))\mathcal{J}^2\Phi)\|^2 &= (\mathcal{J}^2\Phi, a(h(x - \cdot)j_1(\cdot))a^*(h(x - \cdot)j_1(\cdot))\mathcal{J}^2\Phi) \\ &= (\mathcal{J}^2\Phi, [a(h(x - \cdot)j_1(\cdot)), a^*(h(x - \cdot)j_1(\cdot))]\mathcal{J}^2\Phi) + \|a(h(x - \cdot)j_1(\cdot))\mathcal{J}^2\Phi\|^2. \end{aligned} \quad (5.57)$$

The first (commutator) term in (5.58) equals $\int |h(x - y)|^2 j_1(y)^2 dy \|\mathcal{J}^2\Phi\|^2$, which is bounded by $C/D^{2\gamma}$. The second term was treated in (5.55) and is thus bounded by $C/D^{2\gamma}(\Phi, \mathcal{N}\Phi)$. (Note that $(\Phi, \mathcal{N}\Phi) \geq (\mathcal{J}\Phi, \mathcal{N}\mathcal{J}\Phi)$.)

It is immediate that similar estimates hold when aa is replaced by a^*a^* or by a^*a . $\square$

6 Infrared Bounds

In this section we prove Lemma 5.1. It will follow from infrared bounds similar to the ones proved in [1] and [7], which have been proved to hold for

the electrons bound to the Coulomb potential. Those infrared bounds are not sufficient, however, for the localized wave function of the ‘free’ electrons. The chief reason for this insufficiency is that we need to know the dependence of the constants in the infrared bounds on the parameter R_0 . Ultimately, the trouble stems from the fact that we do not know the positions of the n localized electrons, not even remotely. Thus, a direct application of the estimates in [7] would lead to constants that can grow conceivably as R_0^2 or even faster for large R_0 . This problem does not occur for the bound electrons since, in that case, the electrons are localized by the Coulomb potential and the infrared bounds do not depend on the parameter R_0 . The theorem below holds for the localized electrons, as well as for the bound electrons. The proof for bound electrons is easier and is omitted.

As in Sect. 5 the Dirichlet ground state $\Phi_D(n)$ for the ‘free’ electrons localized in $\Omega(B_1, \dots, B_n)$ is denoted simply by Φ . Its energy is $E_D^0(n)$. The ground state for the bound system with Hamiltonian $H^V(N')$ is denoted by Ψ .

LEMMA 6.1 (Infrared bounds). *The following infrared bounds hold for Φ :*

$$\|\widehat{a}_\lambda(k)\Phi\| \leq \frac{C}{|k|^{3/2}} \widehat{\chi}_\Lambda(k) , \quad (6.1)$$

$$\|\widehat{a}_\lambda(k)\Phi\| \leq \frac{CR_0}{|k|^{1/2}} \widehat{\chi}_\Lambda(k) . \quad (6.2)$$

The vector $\widehat{a}_\lambda(k)\Phi$ is a sum of n terms of the form $e^{-ik \cdot Y_j} \widehat{T}_{j,\lambda}(k)$ where $\widehat{T}_{j,\lambda}(k)$ is given by (6.25) and satisfies the estimates

$$\|\nabla_k \widehat{T}_{j,\lambda}(k)\| \leq \frac{CR_0}{|k|^{1/2}(k_1^2 + k_2^2)^{1/2}} \widehat{\chi}_\Lambda(k) + \frac{CR_0}{|k|^{1/2}} \nabla_k \widehat{\chi}_\Lambda(k) \quad (6.3)$$

The vectors Y_j are defined below. The constant C depends on n , on the ultraviolet cutoff Λ and it is a monotone decreasing function of R_0 . Similar bounds but without the factor R_0 hold for the bound electron ground state Ψ .

PROOF: We prove first the bound (6.2) in detail. As a first step one performs an operator valued gauge transformation (see [7, Eq. 47]) by applying the unitary operator (in which A is the vector potential (2.4))

$$U(x) = \exp[-i\sqrt{\alpha} \sum_{j=1}^m \phi_j(x)(x - Y_j) \cdot A(Y_j)] \quad (6.4)$$

to the wave function in each of its variables, i.e.,

$$\Phi \rightarrow \widetilde{\Phi} = \prod_{i=1}^n U(x_i)\Phi =: \mathcal{U}\Phi . \quad (6.5)$$

Here, ϕ_j is a suitably chosen smooth function of compact support and the Y_j are suitably chosen vectors. Note that the factors in the product commute since $A(x)$ commutes with $A(y)$ for all x and y .

Next, we describe the functions ϕ_j . Consider the balls $B_i(2R_0)$, which are concentric with the B_i but with twice the radius, and group them into clusters according to whether they overlap or not. We denote the number of these clusters by m . Such a cluster of balls C_j has a diameter that is bounded above by $4R_0$ times the number of balls in the cluster and hence bounded by $4nR_0$. Denote by Y_j the center of the cluster C_j which is defined to be the center of the smallest ball that contains all the balls $B_i(2R_0)$ belonging to that cluster. We choose functions ϕ_j that are smooth, supported in the union of the balls $B_i(2R_0)$ that belong to the cluster C_j and that are identically one on the union of the balls B_i . In particular $\sum_{j=1}^m \phi_j(x) = 1$ for $x \in \cup_{i=1}^n B_i$.

The gauge transformation $\mathcal{U}$ transforms the Hamiltonian $H^0(n)$ into the Hamiltonian

$$\begin{aligned} \tilde{H}^0(n) = \mathcal{U}H^0(n)\mathcal{U}^* &= \sum_{i=1}^n \left[(p_i + \sqrt{\alpha}\tilde{A}(x_i))^2 + \frac{g}{2}\sqrt{\alpha} \sigma_i \cdot B(x_i) \right] \\ &\quad + \alpha \sum_{i < j} \frac{1}{|x_i - x_j|} + \tilde{H}_f \end{aligned} \quad (6.6)$$

where the new field $\tilde{A}(x) = \mathcal{U}A(x)\mathcal{U}^* + \alpha^{-1/2}\mathcal{U}p\mathcal{U}^* = A(x) + \alpha^{-1/2}\mathcal{U}p\mathcal{U}^*$ is given by

$$\tilde{A}(x) = A(x) - \sum_{j=1}^m \phi_j(x)A(Y_j) - \sum_{j=1}^m \nabla \phi_j(x)(x - Y_j) \cdot A(Y_j), \quad (6.7)$$

in which $A(x)$ is still given by (2.4). The transformed field energy $\tilde{H}_f$ is given as in (2.9) but with $\hat{a}_\lambda(k)$ replaced by the transformed creation and destruction operators

$$b_\lambda(k, X) = \mathcal{U}\hat{a}_\lambda(k)\mathcal{U}^* = \hat{a}_\lambda(k) - i\sqrt{\alpha}w_\lambda(k, X) \quad (6.8)$$

with

$$w_\lambda(k, X) = \sum_{i=1}^n \sum_{j=1}^m \phi_j(x_i)(x_i - Y_j) \cdot \frac{\varepsilon_\lambda(k)}{|k|^{1/2}} e^{-ik \cdot Y_j} \hat{\chi}_\lambda(k). \quad (6.9)$$

As before, the letter X denotes the vector $(x_1, \dots, x_n)$. We note that

$$\hat{a}_\lambda(k)\Phi = \mathcal{U}^* \left[\hat{a}_\lambda(k)\tilde{\Phi} - i\sqrt{\alpha}w_\lambda(k, X)\tilde{\Phi} \right]. \quad (6.10)$$

Since Φ satisfies the Schrödinger equation we can apply the standard pull-through formula [1, 7] and compute

$$\begin{aligned} & \left[\tilde{H}^0(n) - E_D^0(n) \right] \hat{a}_\lambda(k) \tilde{\Phi} = \left[\tilde{H}^0(n), \hat{a}_\lambda(k) \right] \tilde{\Phi} = \\ & 2\hat{\chi}_\Lambda(k) \sqrt{\alpha} |k|^{-1/2} \varepsilon_\lambda(k) \cdot \sum_{i=1}^n (p_i + \sqrt{\alpha} \tilde{A}(x_i)) \left\{ \sum_{j=1}^m (e^{-ik \cdot Y_j} - e^{-ik \cdot x_i}) \phi_j(x_i) \right\} \tilde{\Phi} \\ & + 2\hat{\chi}_\Lambda(k) \sqrt{\alpha} \sum_{i=1}^n \sum_{j=1}^m (p_i + \sqrt{\alpha} \tilde{A}(x_i)) \cdot \nabla \phi_j(x_i) (x_i - Y_j) \cdot \frac{\varepsilon_\lambda(k)}{\sqrt{|k|}} e^{-ik \cdot Y_j} \tilde{\Phi} \\ & + i \frac{g}{2} \hat{\chi}_\Lambda(k) \sqrt{\alpha} \frac{k \wedge \varepsilon_\lambda(k)}{\sqrt{|k|}} \cdot \sum_{i=1}^n \sigma_j e^{-ik \cdot x_i} \tilde{\Phi} - |k| b_\lambda(k, X) \tilde{\Phi} . \quad (6.11) \end{aligned}$$

The term $\sum_{j=1}^m e^{-ik \cdot x_i} \phi_j(x_i)$ stems from the commutator of $\hat{a}_\lambda(k)$ with $A(x_i)$, which yields a term proportional to $e^{-ik \cdot x_i}$ without any summation on j . Using, however, the relation $\sum_{j=1}^m \phi_j(x_i) = 1$, which is valid for any x_i in $\cup_{j=1}^n B_j$, we get that $e^{-ik \cdot x_i} = \sum_{j=1}^m e^{-ik \cdot x_i} \phi_j(x_i)$. Note that x_i has to be in $\cup_{j=1}^n B_j$ for, otherwise, $\tilde{\Phi}$ vanishes.

Likewise, the second term in the above formula appears to have bad infrared behavior, but in order that Φ does not vanish it is necessary that x_i be in one of the balls. But then $\nabla \phi_j(x_i) = 0$ since ϕ_j is constant, and this term does not contribute.

Since $\tilde{H}^0(n) - E_D^0(n)$ is nonnegative the operator $\tilde{H}^0(n) - E_D^0(n) + |k|$ has a bounded inverse $R(k)$. Thus (6.11) leads to the equation

$$\begin{aligned} \hat{a}_\lambda(k) \tilde{\Phi} &= \sqrt{\alpha} \sum_{j=1}^m e^{-ik \cdot Y_j} 2\hat{\chi}_\Lambda(k) |k|^{-1/2} \varepsilon_\lambda(k) \cdot R(k) \\ &\quad \times \sum_{i=1}^n (p_i + \sqrt{\alpha} \tilde{A}(x_i)) (1 - e^{ik \cdot (Y_j - x_i)}) \phi_j(x_i) \tilde{\Phi} \\ &+ i \sqrt{\alpha} \sum_{i=1}^n \frac{g}{2} \hat{\chi}_\Lambda(k) \frac{k \wedge \varepsilon_\lambda(k)}{\sqrt{|k|}} \cdot R(k) \sigma_i e^{-ik \cdot x_i} \tilde{\Phi} + i \sqrt{\alpha} |k| R(k) w_\lambda(k, X) \tilde{\Phi} . \end{aligned} \quad (6.12)$$

As in [7], simple estimates, using Schwarz's inequality, lead to

$$\begin{aligned} \|\widehat{a}_\lambda(k)\widetilde{\Phi}\| &\leq 2\widehat{\chi}_\Lambda(k)\sqrt{\alpha}|k|^{1/2}\|R(k)\sum_{i=1}^n(p_i + \sqrt{\alpha}\widetilde{A}(x_i))^2R(k)\|^{1/2} \\ &\quad \times \left(\sum_{i=1}^n \left(\sum_{j=1}^m \| |x_i - Y_j| \phi_j(x_i) \widetilde{\Phi} \|^2 \right) \right)^{1/2} \\ &\quad + \frac{|g|}{2}\widehat{\chi}_\Lambda(k)\sqrt{\alpha}n|k|^{1/2}\|R(k)\widetilde{\Phi}\| + |k|\|R(k)\|\|w_\lambda(k, X)\widetilde{\Phi}\|. \end{aligned} \quad (6.13)$$

Note that the index j in the first summand is determined by the ball to which the electron i belongs. Therefore, $|x_i - Y_j| \leq 2nR_0$.

Lemma A.1 states that

$$\frac{g\sqrt{\alpha}}{2} \sum_{j=1}^N \sigma_j \cdot B(x_j) + H_f + C \geq 0 \quad (6.14)$$

where

$$C = \frac{1}{8\pi^2} g^2 \alpha N^2 \int \widehat{\chi}_\Lambda(k)^2 dk. \quad (6.15)$$

Thus we also have that

$$\frac{g\sqrt{\alpha}}{2} \sum_{j=1}^N \sigma_j \cdot B(x_j) + \widetilde{H}_f + C \geq 0, \quad (6.16)$$

(note that the gauge transformation $\mathcal{U}$ commutes with $B(x)$) and hence

$$\begin{aligned} \beta &:= \|R(k)\sum_{i=1}^n(p_i + \sqrt{\alpha}\widetilde{A}(x_i))^2R(k)\| \\ &\leq \|R(k)\left(\sum_{i=1}^n[(p_i + \sqrt{\alpha}\widetilde{A}(x_i))^2 + \frac{g}{2}\sigma_i B(x_i)] + \widetilde{H}_f + C\right)R(k)\| \\ &= \|R(k)\left(\widetilde{H}^0(n) + C\right)R(k)\|. \end{aligned} \quad (6.17)$$

Thus, by subtracting and adding $E_D^0 - |k|$,

$$\begin{aligned} \beta &\leq \|R(k)\left(\widetilde{H}^0(n) - E_D^0(n) + |k|\right)R(k)\| + (|E_D^0(n) - |k|| + C)\|R(k)\|^2 \\ &= \|R(k)\| + (|E_D^0(n) - |k|| + C)\|R(k)\|^2 \end{aligned} \quad (6.18)$$

where the constant C depends on Λ and n . For the last term in (6.13) we have that

$$\|w_\lambda(k, X)\widetilde{\Phi}\| \leq \widehat{\chi}_\Lambda(k)|k|^{-1/2} \sum_{j=1}^m \sum_{i=1}^n \| |x_i - Y_j| \phi_j(x_i) \widetilde{\Phi} \|. \quad (6.19)$$

Since $\|R(k)\| = 1/|k|$, we have that $\|R(k)\tilde{\Phi}\| \leq 1/|k|$.

By combining these estimates with (6.13) and using $|x_i - Y_j|\phi_j(x_i) \leq 2nR_0$, we obtain the bound

$$\|\widehat{a}_\lambda(k)\tilde{\Phi}\| \leq CR_0|k|^{-1/2}\|\tilde{\Phi}\|\widehat{\chi}_\lambda(k), \quad (6.20)$$

where the constant C depends on Λ, n and the energy. This estimate carries over to the state Φ by (6.10) since $w_\lambda(k, X)$ applied to Φ satisfies the same estimate, as we see in (6.19). Note that the energy $E_D^0(n)$ does depend on R_0 but it is monotone decreasing as a function of R_0 (by the variational principle for Dirichlet boundary conditions) and it is uniformly bounded below.

Next, we observe in (6.12) that $\widehat{a}_\lambda(k)\tilde{\Phi}$ is a sum of m terms of the form $e^{-ik \cdot Y_j}\widehat{S}_j(k)$ where

$$\begin{aligned} \widehat{S}_{j,\lambda}(k) &= \sqrt{\alpha}2\widehat{\chi}_\Lambda(k)|k|^{-1/2}\varepsilon_\lambda(k) \cdot R(k) \\ &\quad \times \sum_{i=1}^n (p_i + \sqrt{\alpha}\widetilde{A}(x_i))(1 - e^{ik \cdot (Y_j - x_i)})\phi_j(x_i)\tilde{\Phi} \\ &\quad + i\sqrt{\alpha}\frac{g}{2}\widehat{\chi}_\Lambda(k)\frac{k \wedge \varepsilon_\lambda(k)}{\sqrt{|k|}} \cdot R(k) \sum_{i=1}^n \sigma_i\phi_j(x_i)e^{-ik \cdot (x_i - Y_j)}\tilde{\Phi} \\ &\quad + i\sqrt{\alpha}\widehat{\chi}_\Lambda(k)|k|^{1/2}\varepsilon_\lambda(k) \cdot R(k) \sum_{i=1}^n \phi_j(x_i)(x_i - Y_j)\tilde{\Phi}, \end{aligned} \quad (6.21)$$

where we have used the identity

$$\sum_{i=1}^n \sigma_i e^{-ik \cdot x_i} \tilde{\Phi} = \sum_{j=1}^m e^{-ik \cdot Y_j} \sum_{i=1}^n \sigma_i \phi_j(x_i) e^{-ik \cdot (x_i - Y_j)} \tilde{\Phi}. \quad (6.22)$$

Since by (6.8)

$$\widehat{a}_\lambda(k)\Phi = \mathcal{U}^* [\widehat{a}_\lambda(k) - i\sqrt{\alpha}w_\lambda(k, X)] \tilde{\Phi}, \quad (6.23)$$

we obtain that

$$\widehat{a}_\lambda(k)\Phi = \sum_{j=1}^n e^{-ik \cdot Y_j} \widehat{T}_{j,\lambda}(k) \quad (6.24)$$

where

$$\widehat{T}_{j,\lambda}(k) = \mathcal{U}^* [\widehat{S}_{j,\lambda}(k)\tilde{\Phi} - i\sqrt{\alpha}w_\lambda(k, X)] \tilde{\Phi}. \quad (6.25)$$

Differentiating these expressions with respect to k and proceeding in the same fashion as in the proof of (6.2) yields the estimate (6.3).

Differentiating the polarization vectors (2.7) yields the factor $\sqrt{k_1^2 + k_2^2}$ in the denominator of (6.3). The details of the calculation are the same as the ones in [7] and are omitted. The bound (6.1) is considerably easier, since its proof does not require the gauge transformation. Otherwise the proof is word for word as the one above. Finally, the proof of the infrared bounds for Ψ is a word for word translation of the one given in [7]. Note that the localization radius does not show up in this calculation since the electrons are exponentially localized in the vicinity of the origin. $\square$

Finally, we come to the main application of the infrared bounds proved in this section.

PROOF OF LEMMA 5.1: Using (6.1) we can write

$$(\Phi, \mathcal{N}_{\text{out}} \Phi) = \int_{j_2(x) > 0} \|a(x)\Phi\|^2 dx = \int_{j_2(x) > 0} \left\| \sum_{j=1}^n T_{j,\lambda}(x - Y_j) \right\|^2 dx \quad (6.26)$$

which, by Schwarz's inequality is bounded above by

$$\sqrt{n} \sum_{j=1}^n \int_{j_2(x) > 0} \|T_{j,\lambda}(x - Y_j)\|^2 dx. \quad (6.27)$$

For x in the support of j_2 we have that $|x - Y_j| > L$ and hence

$$\int_{j_2(x) > 0} \|T_{j,\lambda}(x - Y_j)\|^2 dx \leq \frac{1}{L^{2\gamma}} \int |x|^{2\gamma} \|T_{j,\lambda}(x)\|^2 dx. \quad (6.28)$$

This last term can be related to the derivative in k space of the function $\widehat{T}_{j,\lambda}(k)$ by the formula

$$\int |x|^{2\gamma} \|T_{j,\lambda}(x)\|^2 dx = C_\gamma \int \frac{(\nabla_k \widehat{T}_{j,\lambda}(k), \nabla_k \widehat{T}_{j,\lambda}(k'))}{|k - k'|^{2\gamma+1}} dk dk' \quad (6.29)$$

where the constant C_γ is given by

$$C_\gamma = (4\pi)^{-3/2} \frac{\Gamma(\frac{1+2\gamma}{2})}{\Gamma(1-\gamma)}. \quad (6.30)$$

Indeed, writing $Q_{j,\lambda}(x) = xT_{j,\lambda}(x)$ this formula follows from

$$\int |x|^{2\gamma-2} \|Q_{j,\lambda}(x)\|^2 dx = C_\gamma \int \frac{(\widehat{Q}_{j,\lambda}(k), \widehat{Q}_{j,\lambda}(k'))}{|k - k'|^{2\gamma+1}} dk dk' \quad (6.31)$$

[10, Corollary 5.10] and the fact that

$$(\widehat{Q}_{j,\lambda}(k), \widehat{Q}_{j,\lambda}(k')) = (\nabla_k \widehat{T}_{j,\lambda}(k), \nabla_k \widehat{T}_{j,\lambda}(k')). \quad (6.32)$$

Using the bound (6.3) a straightforward calculation shows that the function $\|\nabla_k \widehat{T}_{j,\lambda}(k)\|$ is in L^p for all $p < 2$. (The relevant term in (6.3) is the first term on the right side.) Using Schwarz's inequality and the Hardy-Littlewood-Sobolev inequality [10, Theorem 4.3] we can therefore bound (6.29) by

$$C_p \left[\int \|\nabla_k \widehat{T}_{j,\lambda}(k)\|^p dk \right]^{2/p} \leq CR_0^2, \quad (6.33)$$

with $p = 6/(5 - 2\gamma)$, which is strictly less than 2 for $\gamma < 1$. To prove (5.3) we write for some $0 < H$

$$(\Phi, \mathcal{N}\Phi) = \sum_{\lambda} \int_{|k| \leq H} \|\widehat{a}_{\lambda}(k)\Phi\|^2 dk + \sum_{\lambda} \int_{H \leq |k|} \|\widehat{a}_{\lambda}(k)\Phi\|^2 dk, \quad (6.34)$$

and, using (6.2) and (6.1) we get

$$CR_0 \int_{|k| \leq H} \frac{1}{|k|} \widehat{\chi}_{\Lambda} dk + C \int_{H \leq |k|} \frac{1}{|k|^3} \widehat{\chi}_{\Lambda} dk, \quad (6.35)$$

which, optimized over H , leads to (5.3).

The proof for the state Ψ is carried out in precisely the same fashion. $\square$

Appendix A

LEMMA A.1. *On $\wedge^N L^2(\mathbb{R}^3; \mathbb{C}^2) \otimes \mathcal{F}$ we have that*

$$\frac{g\sqrt{\alpha}}{2} \sum_{j=1}^N \sigma_j \cdot B(x_j) + H_f + \frac{1}{8\pi^2} g^2 \alpha N^2 \int \widehat{\chi}_{\Lambda}(k)^2 dk \geq 0. \quad (A.1)$$

PROOF: The magnetic field operator can be written in the form

$$\sum_{\lambda=1}^2 \int [c_{\lambda}^*(k) \widehat{a}_{\lambda}(k) + c_{\lambda}(k) \widehat{a}_{\lambda}^*(k)] dk \quad (A.2)$$

where

$$c_{\lambda}^*(k) = \frac{g\sqrt{\alpha}}{4\pi} \widehat{\chi}_{\Lambda}(k) \sum_j \frac{ik \wedge \varepsilon_{\lambda}(k)}{\sqrt{|k|}} \cdot \sigma_j e^{ik \cdot x_j}. \quad (A.3)$$

With this notation we can write the Hamiltonian (A.1) as

$$\begin{aligned} & \sum_{\lambda=1}^2 \int |k| \left[\widehat{a}_{\lambda}^*(k) \otimes I + \frac{1}{|k|} c_{\lambda}^*(k) \right] \left[\widehat{a}_{\lambda}(k) \otimes I + \frac{1}{|k|} c_{\lambda}(k) \right] dk \\ & - \sum_{\lambda=1}^2 \int \frac{1}{|k|} c_{\lambda}^*(k) c_{\lambda}(k) dk \end{aligned} \quad (A.4)$$

Here, I denotes the identity operator on spin space. The first term is non-negative and crude estimates on the second yield the lemma. $\square$

Appendix B

In this section we prove the estimates on the coupling functions $h_\lambda^i(y)$ which are defined by

$$h_\lambda^i(y) = \frac{1}{2\pi} \int \frac{1}{\sqrt{|k|}} \varepsilon_\lambda^i(k) \chi_\Lambda(k) e^{ik \cdot y} dk. \quad (\text{B.1})$$

It is important to choose the polarization vectors carefully in order that their Fourier transforms (from k -space to y -space) have nice decay properties as $|y|$ tends to infinity. We shall express these decay properties in an integrated form. The reason for that is that the decay is not uniform with respect to the direction of the y variable. Recall the definitions (2.7).

LEMMA B.1 (Decay of the coupling functions). *For any $\gamma < 1$ there is a finite constant $C(\gamma)$ such that*

$$\sum_{i=1}^3 \sum_{\lambda=1}^2 \int |y|^{2\gamma} |h_\lambda^i(y)|^2 dy \leq C(\gamma). \quad (\text{B.2})$$

PROOF: First we compute the gradient of h_λ^i in k -space. It is elementary that

$$\left| \nabla \frac{1}{\sqrt{|k|}} \varepsilon_\lambda(k) \right| \leq \frac{C}{\sqrt{|k|} \sqrt{k_1^2 + k_2^2}}, \quad (\text{B.3})$$

where C is some constant. Because χ_Λ is smooth,

$$\sum_{i=1}^3 \sum_{\lambda=1}^2 \int \left| \nabla \frac{1}{\sqrt{|k|}} \varepsilon_\lambda(k) \chi_\Lambda(k) \right|^p dk \leq C(p) \quad (\text{B.4})$$

for any $p < 2$. We proceed as in (6.29), (6.30) and write

$$\sum_{i=1}^3 \sum_{\lambda=1}^2 \int |y|^{2\gamma} |h_\lambda^i(y)|^2 dy = C_\gamma \sum_{i=1}^3 \sum_{\lambda=1}^2 \int \frac{\nabla \widehat{h}_\lambda^i(k) \cdot \nabla \widehat{h}_\lambda^i(k')}{|k - k'|^{2\gamma+1}} dk dk'. \quad (\text{B.5})$$

Again, by the Hardy-Littlewood-Sobolev inequality [10, Theorem 4.3] this is bounded by

$$C_p \left[\sum_{i=1}^3 \sum_{\lambda=1}^2 \int \left| \nabla \frac{1}{\sqrt{|k|}} \varepsilon_\lambda(k) \chi_\Lambda(k) \right|^p dk \right]^{2/p} \leq C_p C(p)^{2/p}, \quad (\text{B.6})$$

where $p = 6/(5 - 2\gamma) < 2$ if $\gamma < 1$. $\square$

References

- [1] V. Bach, J. Fröhlich and I. Sigal, *Spectral analysis for systems of atoms and molecules coupled to the quantized radiation field*, Comm. Math. Phys. **207**, 249–290 (1999).
- [2] V. Bach, J. Fröhlich and I. Sigal, *Quantum electrodynamics of confined nonrelativistic particles*, Adv. Math. **137**, 299–395 (1998) .
- [3] J-M. Barbaroux, T. Chen and S. Vugalter, *Binding conditions for atomic N -electron systems in non-relativistic QED* arXiv math-ph/0304019.
- [4] T. Chen *Operator-theoretic infrared renormalization and construction of dressed 1-particle states in non-relativistic QED*, arXiv math-ph/0108021, slightly modified preprint version of ETH-Diss 14203.
- [5] C. Gérard, *On the existence of ground states for massless Pauli-Fierz Hamiltonians* Ann. Henri Poincaré **1**, 443–459 (2000).
- [6] M. Griesemer, *Exponential decay and ionization thresholds in non-relativistic quantum electrodynamics*, arXiv math-ph/0206024.
- [7] M. Griesemer, E. H. Lieb and M. Loss, *Ground states in non-relativistic quantum electrodynamics*, Invent. Math. **145**, 557–595 (2001).
- [8] M., Hirokawa *An expression of the ground state energy of the spin-boson model* J. Funct. Anal. **162**, 178–218 (1999).
- [9] F. Hiroshima, *Self-adjointness of the Pauli-Fierz Hamiltonian for arbitrary coupling constants*, Ann. Henri Poincaré **3**, 171–201 (2002).
- [10] E.H. Lieb and M. Loss, *Analysis*, Amer. Math. Soc. second edition, 2001.
- [11] E. H. Lieb *Proofs of some conjectures on permanents* J. Math. and Mech. **16**, 127–139 (1966).
- [12] E.H. Lieb, J. Fröhlich and M. Loss, *Stability of Coulomb Systems with Magnetic Fields I. The One-Electron Atom*, Commun. Math. Phys. **104**, 251–270 (1986).
- [13] H. Spohn, *Ground state of a quantum particle coupled to a scalar Bose field* Lett. Math. Phys. **44**, 9–16 (1998) .
- [14] G. Zhislin, *A study of the spectrum of the Schrödinger operator for a system of several particles*, Trudy Moscov Mat. Obsc. **9**, 81–120 (1960).

A Bound on Binding Energies and Mass Renormalization in Models of Quantum Electrodynamics

Elliott H. Lieb¹ and Michael Loss²

Received October 22, 2001; accepted January 7, 2002

We study three models of matter coupled to the ultraviolet cutoff, quantized radiation field and to the Coulomb potential of arbitrarily many nuclei. Two are nonrelativistic: the first uses the kinetic energy $(p + eA(x))^2$ and the second uses the Pauli–Fierz energy $(p + eA(x))^2 + e\sigma \cdot B(x)$. The third, no-pair model, is relativistic and replaces the kinetic energy with the Dirac operator $D(A)$, but restricted to its positive spectral subspace, which is the “electron subspace.” In each case we are able to give an upper bound to the *binding* energy—as distinct from the less difficult ground state energy. This implies, for the first time we believe, an estimate, albeit a crude one, of the mass renormalization in these theories.

KEY WORDS: Quantum electrodynamics; mass renormalization.

1. INTRODUCTION

There has been a great deal of recent work dedicated to the construction of a theory of ordinary bulk matter interacting with the quantized radiation field. In such theories the number of electrons, N , is usually held fixed (i.e., pair production is not allowed) and these N particles interact with each other and with K fixed nuclei via the ordinary electrostatic Coulomb potential—in the Coulomb gauge. (The nuclei are fixed because they are, relative to the electrons, infinitely massive.) The electrons also interact with

Dedicated to two masters of mathematical physics, David Ruelle and Yakov Sinai, on the occasion of their 65th birthdays.

¹ Departments of Mathematics and Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544; e-mail: lieb@math.princeton.edu

² School of Mathematics, Georgia Tech, Atlanta, Georgia 30332; e-mail: loss@math.gatech.edu

the magnetic vector potential A which is *quantized* and which has the well known quantized field energy. It is essential, however, to have an ultra-violet cutoff A in the interaction of electrons and A field, for otherwise almost everything in the theory becomes infinite. This is not an enormous drawback since we are interested only in the low energy physics of atoms, molecules and photons.

Attention has been paid mostly to the stability of matter, namely that the ground state energy is bounded below by $C(N+K)$, where C is some universal constant (depending on the parameters of the theory but not on N and K).^(1–6) Almost no attention has been paid to the estimation of the atomic binding energy, i.e., to the difference between the ground state energies with and without the Coulomb potential.

In this paper we shall consider two nonrelativistic theories and one relativistic theory. In the absence of quantized fields, the ground state energy (or “self energy”) of a free electron (i.e., without other electrons and nuclei) is zero in the nonrelativistic case and equals mc^2 in the relativistic case. (Here, m is the unrenormalized, or “bare” mass of the electron.) When N electrons are present, but without nuclei, the energy is still zero (or Nmc^2) because the electrons can move infinitely far apart. Therefore, in the presence of nuclei, the ground state energy is equal to the binding energy (or equals the binding energy plus Nmc^2) when there is no quantization of the A field.

The situation changes dramatically when the A field is quantized. The self-energy of a free electron (i.e., the ground state energy without Coulomb potentials but with the quantized A field) is large if A is not too small. If the fine structure constant $\alpha = e^2/\hbar c$ is not too large (e.g., $1/137$) and if the nuclear charges are not too large then the change in the ground state energy is not very large. Thus, the binding energy is the difference of two large quantities and its calculation is like “looking for a needle in a haystack.”

All three models use “minimal coupling,” i.e., p is replaced by $p + eA(x)/c$ in the kinetic energy. The first has no explicit spin interaction with the magnetic field while the second—the Pauli–Fierz model—has a $\sqrt{\alpha} \sigma \cdot B(x)$ term. The latter is much more delicate than the former (e.g., it requires bounds on α and Z , as well as the presence of a field energy, for stability, while the former needs no such restrictions), and our results in the second case are not as good as in the first. Still, they are meaningful. The third theory uses the Dirac operator and is relativistic (except for an ultraviolet cutoff in the A field).

Clearly, it is essential to have a decent grasp of the binding energy, which is the truly physical quantity, in order to be able to proceed with a non-perturbative renormalization program. It is useful to recall that the

binding energy for one electron and one nucleus of charge Z should be $\frac{1}{2} m_{\text{phys}} c^2 \alpha^2 Z^2$, where m_{phys} is the physical, renormalized electron mass. In the case of hydrogen, $Z = 1$, this is a comparatively tiny energy. (There is no charge renormalization in a theory, without pair production, but there is one in a theory with pair production, in which case α here must be replaced by the renormalized α .)

Our purpose here is to find an upper bound to the binding energy in the two non-relativistic QED theories that have been extensively studied in the literature and to the relativistic theory in ref. 3. We believe it is the first time that such a rigorous estimate has been made. Note that the binding energy (14) is always positive, by definition, and thus an upper bound corresponds to a lower bound on the energy of the fully interacting system relative to the energy of free electrons.

If we equate the binding energy (which depends on the unrenormalized, or “bare” mass m) to the physical binding energy (which depends on m_{phys}) we obtain a bound to the amount of renormalization that is needed (see (23), (27), and (29–31) for the first model). We believe that these are the first rigorous renormalization estimates of their kind in a (admittedly simple-minded) quantized theory of electrons and photons.

Let us note an interesting feature of our results about the mass. There are several ways to define the renormalized mass. The usual one is to look at the ground state energy of a free electron with *fixed* total momentum (electron + field) and to define $1/2m_{\text{phys}}$ to be the coefficient of p^2 in the energy at $p = 0$. Another way is to set the binding energy equal to the physical binding energy, as is done here.

- The latter definition has the property that for every value of Λ and of m_{phys} there is a value of m that gives equality. (Our bound is unique, but the true answer is, conceivably, not unique.)

- The former, usual definition very likely yields a solution for m only if Λ and $1/m_{\text{phys}}$ are sufficiently small. (We always assume that $m \geq 0$ in order that the notion of a ground state makes sense for the unrenormalized theory.)

We cannot prove this last statement but it agrees with the prediction of perturbation theory and with classical electromagnetism. It also agrees with Van Kampen’s exact solution of Kramer’s “dipole approximation” model,⁽⁷⁾ which was the model that gave impetus to the renormalization program in QED. Instead of $(p + A(x))^2$ as in our first model, one takes $(p + A(0))^2$. In such a model the energy of N free electrons goes as $CA^{3/2} \sqrt{N}$ instead of $CA^{3/2}N$, as in our model (see (21)), but this is not the most significant point. For $N = 1$ one can compute the energy as a

function of momentum p and finds that, with the first definition, $m_{\text{phys}} = C'm + \alpha A$. Consequently, there is no solution for m if A or $1/m_{\text{phys}}$ are large.

The first Hamiltonian H_N we consider is given by

$$\frac{1}{\hbar c} H_N = \mathcal{T} + \alpha V_c + H_f, \quad (1)$$

where $\mathcal{T}$ is proportional to the kinetic energy of N electrons

$$\mathcal{T} = \sum_{j=1}^N T_j \quad (2)$$

with

$$T_j = \frac{\hbar}{2mc} (p_j + \sqrt{\alpha} A(x_j))^2, \quad (3)$$

with $p_j = i\nabla_{x_j}$, and where m is the (unrenormalized or bare) mass of the electron. The quantized, ultraviolet cutoff electromagnetic vector potential is

$$A(x) = \frac{1}{2\pi} \sum_{\lambda=1}^2 \int_{|k| \leq A} \frac{\varepsilon_{\lambda}(k)}{\sqrt{\omega(k)}} (a_{\lambda}(k) e^{ik \cdot x} + a_{\lambda}^*(k) e^{-ik \cdot x}) dk, \quad (4)$$

where A is the ultraviolet cutoff on the wave-numbers $|k|$. The operators $a_{\lambda}, a_{\lambda}^*$ satisfy the usual commutation relations

$$[a_{\lambda}(k), a_{\nu}^*(q)] = \delta(k-q) \delta_{\lambda,\nu}, \quad [a_{\lambda}(k), a_{\nu}(q)] = 0, \quad \text{etc.} \quad (5)$$

and the vectors $\varepsilon_{\lambda}(k)$ are the two possible orthonormal polarization vectors perpendicular to k and to each other. The field energy is

$$H_f = \sum_{\lambda=1,2} \int_{\mathbb{R}^3} \omega(k) a_{\lambda}^*(k) a_{\lambda}(k) dk; \quad (6)$$

the physical choice of ω is $\omega(k) = |k|$, but our Theorems 2.1 and 3.1 are not restricted to this choice. No infrared cutoff is needed.

Finally, there is the Coulomb potential. There are K nuclei with positive charges $eZ_1, \dots, eZ_K$ and with fixed locations $R_1, \dots, R_K$ in $\mathbb{R}^3$. In this model the nuclei will preferentially locate themselves at those R_j that minimize the total energy, but these special locations are irrelevant for our theorem.

$$V_c = - \sum_{i=1}^N \sum_{k=1}^K Z_k \frac{1}{|x_i - R_k|} + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|} + \sum_{1 \leq k < l \leq K} Z_k Z_l \frac{1}{|R_k - R_l|}. \quad (7)$$

The Hilbert space is $\mathcal{H}_N = \bigwedge_{i=1}^N L^2(\mathbb{R}^3; C^q) \otimes \mathcal{F}$, where $\bigwedge$ denotes the antisymmetric tensor product (Fermi statistics), $\mathcal{F}$ is the photon Fock space, q is the number of spin states for each electron ($q = 2$ in nature). The Hamiltonian (1) is bounded below by $C(N+K)$ (even without the aid of H_f); this was first stated in ref. 8, p. 857 and ref. 9, p. 2, who noted that the stability proof in ref. 10 with $A = 0$ extends to the $A \neq 0$ case by virtue of a diamagnetic inequality. In ref. 8 a remark of J. Fröhlich is presented, that this applies equally well to the quantized A field (4) because $[A(x), A(y)] = 0$ for all x, y . We use and discuss this fact again in (17) later.

The Pauli–Fierz Hamiltonian, H_N^{PF} , which is treated in Section 3, is

$$H_N^{\text{PF}} = H_N + \frac{\hbar^2}{2m} \sqrt{\alpha} \sum_{j=1}^N \sigma_j \cdot B(x_j) \quad \text{with} \quad B(x) = \text{curl } A(x). \quad (8)$$

The Pauli matrices σ_j are the spin- $\frac{1}{2}$ operators for particle j .

The third model is the relativistic no-pair model treated in ref. 3, whose Hamiltonian is

$$\frac{1}{\hbar c} H_N^{\text{rel}} = P^+ \left\{ \sum_{i=1}^N D_i(A) + \alpha V_c + H_f \right\} P^+. \quad (9)$$

Here, $D(A)$ is the Dirac operator

$$D(A) := \vec{\alpha} \cdot (-i\nabla + \sqrt{\alpha} A(x)) + \frac{mc}{\hbar} \beta, \quad (10)$$

and P^+ is the projector onto the positive spectral subspace of $D(A)$ for all of the N electrons. Since the N Dirac operators commute with each other, this definition of P^+ as a projector makes sense. In other words, we start with the usual Hilbert space $\mathcal{H}_N$ and then replace it by the smaller physical Hilbert space $\mathcal{H}_N^{\text{phys}} = P^+ \mathcal{H}_N$. In $\mathcal{H}_N^{\text{phys}}$ it is impossible to separate the L^2 spaces from the Fock space.

While energy, being one component of a four-vector, is not a relativistically invariant quantity, it is true, nevertheless, that positive and negative energies of $D(A)$ are relativistic concepts since they are invariant under Lorentz transformations that do not change the direction of time. We thank J.-M. Graf for this remark and we thank J. Yngvason for noting that for this to be true it is essential that the joint spectrum of energy and momentum of $D(A)$ lies in the light cone. We have not proved this, but it is plausibly true.

2. BINDING ENERGY BOUND (SIMPLE VERSION)

In this section we analyze the binding energy for H_N . We define $Z = \max\{Z_1, \dots, Z_K\} \geq 0$ and set

$$\kappa = \frac{\pi}{2} Z + 2.2159q^{1/3}Z^{2/3} + 1.0307q^{1/3}. \quad (11)$$

We also define the (positive) free-electron ground state energy

$$E_0(N) = \inf \text{spec } H_N(0) > 0 \quad (12)$$

(where $H_N(0)$ is the Hamiltonian (1) without the V_c term), the total ground state energy

$$E(N) = \inf \text{spec } H_N \leq E_0(N), \quad (13)$$

and the (positive) binding energy

$$\Delta E(N) = E_0(N) - E(N). \quad (14)$$

Theorem 2.1 (Binding Energy for H_N). Assume one of two cases:

$$E_0(N) \begin{cases} \leq \frac{1}{2} (\kappa\alpha)^2 mc^2 N, & \text{Case A} \\ \geq \frac{1}{2} (\kappa\alpha)^2 mc^2 N, & \text{Case B.} \end{cases} \quad (15)$$

Then the binding energy per electron satisfies

$$\frac{\Delta E(N)}{N} \leq \begin{cases} (\kappa\alpha)^2 mc^2 & \text{Case A} \\ \kappa\alpha \sqrt{2mc^2} \sqrt{E_0(N)/N} & \text{Case B.} \end{cases} \quad (16)$$

Proof. We use the known result for the stability of “relativistic” matter in the form given in ref. 11, Eqs. (2.9) and (5.2) (which improves some of the results in ref. 6).

$$V_c \geq -\kappa \sum_{j=1}^N |p_j + \sqrt{\alpha} A(x_j)| \quad (17)$$

for any vector field $A(x)$. (Note that although (17) was proved for ordinary, non-quantized A fields, we are allowed to use it for our operator-valued field (4) since the commutator $[A(x), A(y)] = 0$ for all x, y , and hence there is a representation in which $A(x)$ is an ordinary vector field. Of course, $A(x)$ does not commute with H_f but that is immaterial.)

By Schwarz's inequality

$$V_c \geq -\kappa \sqrt{2mc/\hbar} \sqrt{N} \sqrt{\mathcal{T}}. \quad (18)$$

Since $H_f \geq 0$ and since the square root is operator monotone,

$$\begin{aligned} \frac{1}{\hbar c} H_N &\geq \mathcal{T} + H_f - \kappa \alpha \sqrt{2mc/\hbar} \sqrt{N} \sqrt{\mathcal{T}} \\ &\geq \mathcal{T} + H_f - \kappa \alpha \sqrt{2mc/\hbar} \sqrt{N} \sqrt{\mathcal{T} + H_f}. \end{aligned} \quad (19)$$

The final step is to note that the function $f(x) = x - \kappa \alpha \sqrt{2mc/\hbar} \sqrt{N} \sqrt{x}$ has its minimum at $x = \kappa^2 \alpha^2 mcN/2\hbar$. Therefore, when $x \leq \kappa^2 \alpha^2 mcN/2\hbar$ we can say that $f(x)$ is not less than the minimum of $f(x)$, which is $-\frac{1}{2}(\kappa \alpha)^2 mcN/\hbar$. This is Case A of (15). Therefore $\Delta E \leq E_0(N) + \frac{1}{2}(\kappa \alpha)^2 mc^2 N \leq (\kappa \alpha)^2 mc^2 N$, which is Case A of (16). On the other hand, if $x \geq \kappa^2 \alpha^2 mcN/2\hbar$ we can say that $f(x)$ is monotone increasing in x . Since $x = \mathcal{T} + H_f \geq E_0(N)/\hbar c \geq \kappa^2 \alpha^2 mcN/2\hbar$, the infimum of the spectrum of the right side of (19) is given by $f(E_0(N)/\hbar c)$, which is Case B of (16). ■

To apply this Theorem 2.1 we must have a decent estimate of $E_0(N)$. Let us consider the physical case $\omega(k) = |k|$ and let us define (with $\lambda_C = \hbar c/m_{\text{phys}} = \text{physical Compton wavelength}$)

$$\mathcal{R} = \frac{\hbar c}{mc^2} A = \frac{m_{\text{phys}}}{m} A \lambda_C, \quad (20)$$

which is the ratio of the cutoff photon energy to the self energy that an electron would have in a relativistic theory. A bound on $E_0(N)$ in this case is provided in ref. 12 where a proof is announced and outlined that for *fermions* there are constants C_1, C_2 (depending on q) such that (for large A and fixed α)

$$C_1 mc^2 \alpha^{1/2} \mathcal{R}^{3/2} N \leq E_0(N) \leq C_2 mc^2 \alpha^{2/7} \mathcal{R}^{12/7} N \quad (21)$$

The exact exponent is still not known but we lean towards 12/7. In any case, it differs from the perturbation theory value 2. Fermions are most important here because one can show⁽¹²⁾ that $C_3 mc^2 \alpha^{1/2} \mathcal{R}^{3/2} N^{1/2} < E_0(N) < C_4 mc^2 \alpha^{2/7} \mathcal{R}^{12/7} N^{5/7}$ for bosons, and this would be useless for our purposes. Unfortunately, the bounds in (21) do not imply that $E_0(N)$ is strictly linear in N , as one would hope. We also note that (21) holds even if the Coulomb repulsion among the electrons is omitted.

If $\mathcal{R}$ is small then we are in Case A. This will surely be so if $\kappa^2 \alpha^{12/7} \geq 2C_2 \mathcal{R}^{12/7}$. Let us note that $\kappa \leq 5.67Z$ for $Z \geq 1$ and $q = 2$. Then,

$$\frac{\Delta E(N)}{N} \leq (32.2) \alpha^2 Z^2 m c^2 \quad \text{Case A,} \quad (22)$$

which compares not unfavorably with the hydrogenic value $e_{\text{hydrogenic}}(Z) = Z^2 \alpha^2 m_{\text{phys}} c^2 / 2$, where m_{phys} is the physical electron mass. As $\mathcal{A}$ increases the bare mass m should decrease. If we set $\Delta E(N)/N = e_{\text{hydrogenic}}(Z)$, inequality (22) tells us that the required bare mass cannot be too small, namely

$$m \geq m_{\text{phys}} / 64.4 \quad \text{Case A.} \quad (23)$$

We turn now to Case B, which surely holds if

$$\kappa^2 \alpha^{3/2} \leq 2C_1 \mathcal{R}^{3/2}. \quad (24)$$

With the help of (21) we can conclude that

$$\frac{\Delta E(N)}{N} \leq \sqrt{2C_2} \kappa \alpha^{8/7} \mathcal{R}^{6/7} m c^2 \quad \text{Case B.} \quad (25)$$

To understand Case B further, let us use $\kappa \leq 5.67Z$ and note that (25) becomes

$$\frac{\Delta E(N)}{N e_{\text{hydrogenic}}(Z)} \leq \frac{5.67}{Z} \frac{C_2^{1/2}}{C_1^{4/7}} 2^{13/14} \mu^{8/7} \frac{m}{m_{\text{phys}}} \quad \text{if } 5.67Z \leq \mu, \quad (26)$$

with $\mu = \alpha^{-3/4} \sqrt{2C_1} \mathcal{R}^{3/4}$.

To satisfy the condition in (26) for all $Z \leq 92$ we can take $\mu = 5.67 \cdot 92 = 522$ or $\mathcal{R} = (522)^{4/3} (2C_1)^{-2/3} \alpha = 30.7 (2C_1)^{-2/3}$ (with $\alpha = 1/137$). This means that we fix $\mathcal{A}$ in units of the *bare* Compton wavelength $\hbar c/m$. Assuming that we choose C_1 to be not too large (which can always be done since C_1 refers to a lower bound in (21)), this allows for a sizeable value of the cutoff $\mathcal{A}$ (see (20)). Now let us set the left side of (26) equal to 1, in order to make contact with experiment. We then find (since $Z \geq 1$) that

$$m \geq m_{\text{phys}} C_1^{4/7} C_2^{-1/2} / 13,800 \quad \text{Both Cases.} \quad (27)$$

Alternatively, we may measure $\mathcal{A}$ in terms of the *physical* Compton wavelength λ_c . That the bare mass cannot be too small can be seen as follows. Consider the following inequality, which is related to (24)

$$(5.67 \cdot 92)^2 \alpha^{3/2} \leq 2C_1 \mathcal{R}^{3/2}. \quad (28)$$

If this inequality fails then we have the bound

$$m \geq m_{\text{phys}} \mathcal{A} \lambda_c (2C_1)^{2/3} (5.67 \cdot 92)^{-4/3} \alpha^{-1} = 1.74 \cdot 10^{-6} (2C_1)^{2/3} \mathcal{A} \lambda_c m_{\text{phys}} \quad (29)$$

with $\alpha = 1/137$.

On the other hand if (28) holds then (24) holds for all $Z \leq 92$ (since $\kappa \leq 5.67Z$). Then we are in Case B and if we express the right of (26) entirely in terms of $\mathcal{A} \lambda_c$ we find (for all $Z \leq 92$)

$$m \geq (2 \cdot 5.67 \sqrt{2C_2})^{-7} \alpha^6 (92)^7 (\mathcal{A} \lambda_c)^{-6} m_{\text{phys}} \quad (30)$$

$$= 3.0 \times 10^{-8} C_2^{-7/2} (\mathcal{A} \lambda_c)^{-6} m_{\text{phys}}. \quad (31)$$

3. BINDING ENERGY BOUND (PAULI-FIERZ VERSION)

In this section we analyze the binding energy for H_N^{PF} in (8). The ground state energy and binding energy are defined as before in (12), (13), and (14), but with H_N^{PF} , and we do not encumber the notation with a superscript PF. As far as constants are concerned, the following theorem is not the best possible one, but it is presented this way for simplicity. In particular, we do not have to assume that α is bounded—as we do in the hypothesis of Theorem 3.1.

Some constants have to be defined. The maximum nuclear charge Z is defined as before and we then define $\Omega = \max\{Z, 20.6\}$. (Note, for later use, that $20.6 = 64.5/\pi$.) We also define $\xi = (0.060)(8\pi)(3/4)^{-3/2} = 2.322$ and

$$\Gamma = 9.65 \left(\frac{\alpha K}{N} \right)^{1/4} \mathcal{A}. \quad (32)$$

We also define the operator $\hbar c \mathcal{T}'$ to be the total Pauli–Fierz kinetic energy, namely,

$$\mathcal{T}' = \sum_{j=1}^N T_j^{\text{PF}} = \mathcal{T} + \frac{\hbar}{2mc} \sqrt{\alpha} \sum_{j=1}^N \sigma_j \cdot B(x_j) \geq 0, \quad (33)$$

where

$$\begin{aligned} T_j^{\text{PF}} &= \frac{\hbar}{2mc} \{ (p_j + \sqrt{\alpha} A(x_j))^2 + \sqrt{\alpha} \sigma_j \cdot B(x_j) \} \\ &= \frac{\hbar}{2mc} \{ \sigma_j \cdot (p_j + \sqrt{\alpha} A(x_j)) \}^2. \end{aligned} \quad (34)$$

Theorem 3.1 (Binding Energy for H_N^{PF}). Assume that $2\pi\Omega\xi\alpha^2 < 1$ and assume one of two cases:

$$E_0(N) \begin{cases} \leq 2(1 - 2\pi\Omega\xi\alpha^2)^{-2} (\pi\Omega\alpha)^2 mc^2 N, & \text{Case A} \\ \geq 2(1 - 2\pi\Omega\xi\alpha^2)^{-2} (\pi\Omega\alpha)^2 mc^2 N, & \text{Case B.} \end{cases} \quad (35)$$

Then the binding energy per electron satisfies

$$\frac{\Delta E(N)}{N} \leq \begin{cases} 2(\pi\Omega\alpha)^2 (1 - 2\pi\Omega\xi\alpha^2)^{-2} (2 - 2\pi\Omega\xi\alpha^2) mc^2 + 2\pi\Omega\alpha\Gamma\hbar c & \text{Case A} \\ 2\pi\Omega\xi\alpha^2 E_0(N)/N + 2\pi\Omega\alpha\sqrt{2mc^2}\sqrt{E_0(N)/N} + 2\pi\Omega\alpha\Gamma\hbar c & \text{Case B.} \end{cases} \quad (36)$$

Proof. The strategy is the same as in Theorem 2.1. An analogue of (17) is provided by ref. 3.

$$\frac{1}{2\pi\Omega} V_c \geq - \sum_{j=1}^N |\sigma_j \cdot (p_j + \sqrt{\alpha} A(x_j))| - \xi\alpha H_f - \Gamma N, \quad (37)$$

The derivation of (37) from ref. 3 will be explained at the end of the proof. For the moment let us continue with the proof of the theorem. The analogue of (18) is then

$$\frac{1}{2\pi\Omega} V_c \geq -\sqrt{2mc/\hbar}\sqrt{N}\sqrt{\mathcal{T}'} - \xi\alpha H_f - \Gamma N. \quad (38)$$

Consequently, since $\mathcal{T}' \geq 0$,

$$\begin{aligned} \frac{1}{\hbar c} H_N &\geq (\mathcal{T}' + H_f)(1 - 2\pi\Omega\xi\alpha^2) \\ &\quad - 2\pi\Omega\alpha\sqrt{2mc/\hbar}\sqrt{N}\sqrt{\mathcal{T}' + H_f} - 2\pi\Omega\alpha\Gamma N. \end{aligned} \quad (39)$$

The rest follows as in the proof of Theorem 2.1.

It remains to show how (37) arises from ref. 3. We consider a Hamiltonian, $\hat{H}_N$, similar to H_N'' in ref. 3, but with some auxiliary parameters.

$$\hat{H}_N = \sum_{j=1}^N |\sigma_j \cdot (p_j + \sqrt{\alpha} A(x_j))| + \alpha' V_c + \gamma H_f, \quad (40)$$

with $\alpha', \gamma > 0$. Note that $\hat{H}_N$ has the dimension of length⁻¹ and not an energy. Note also that the α appearing in $p_j + \sqrt{\alpha} A(x_j)$ is the true α and not α' .

We shall prove the analogue of Theorem 2.2 of ref. 3 with $\hat{H}_N$ in place of H''_N and with an appropriate substitute for the lower bound stated in Theorem 2.2. Theorem 3.1 of ref. 3 is unchanged, and we turn to Section 4. We take $\varepsilon = 0$ (it can be taken to be zero even if $m \neq 0$). All the equations in this section remain true if we replace $\kappa\alpha$ by $\kappa\alpha'$ and $\kappa^2\alpha^2$ by $\kappa^2(\alpha')^2$. Thus, in the lower bound for the operator H_2 in Section 4 of ref. 3, there is α in the numerator and α' in the denominator.

The next step is to use the inequality in Example 1 of Appendix B of ref. 3 to bound $\int B^2$ appearing in the lower bound for H_2 , but we have to remember that we have γH_f and not H_f . Thus, the analogous conditions on the parameters are $\kappa \leq \pi\Omega$, $\kappa\alpha' < 1$ and $\gamma \geq 8\pi(0.060)\alpha(1 - \kappa^2(\alpha')^2)^{-3/2}$.

We make the choices

$$\kappa = \pi\Omega, \quad \alpha' = (2\pi\Omega)^{-1}, \quad \gamma = \xi\alpha. \quad (41)$$

The lower bound to $\hat{H}_N$ is as in ref. 3. We find that the corresponding number C_2 is bounded by

$$C_2^4 = \frac{N}{K} \left[\frac{6 + (\alpha'/2)(\sqrt{2Z} + 2.3)^2}{(27/2\pi)\gamma} \right] < \frac{N}{K} \left[\frac{13\pi}{27\gamma} \right], \quad (42)$$

since $(\sqrt{2Z} + 2.3)^2/2\pi\Omega < 1$.

The analogous lower bound for the operator $\hat{H}_N$, is then

$$\frac{1}{N} \hat{H}_N > -\frac{K}{N} \left[\frac{18 A\gamma}{\pi} \right] C_2^3 > -\left(\frac{\alpha K}{N} \right)^{1/4} A\nu \quad (43)$$

with $\nu = (\xi/\pi)^{1/4} (2/3)(39)^{3/4} = 9.65$.

This completes the derivation of (37) from ref. 3. ■

Using Theorem 3.1 we could proceed to derive more explicit bounds for the binding energy and m_{phys} —as in the discussion after 2.1, but we leave this task to the interested reader. The only needed information is the analogue of (21) for the Pauli–Fierz operator. As announced in ref. 12

$$C_1 mc^2 \frac{\alpha}{1 + \alpha^{2/3}} \mathcal{R}N \leq E_0(N) \leq C_2 mc^2 \alpha^{1/2} \mathcal{R}^{3/2} N. \quad (44)$$

4. BINDING ENERGY BOUND (RELATIVISTIC VERSION)

We consider the Hamiltonian H_N^{rel} in (9). Our results here are very crude and we state them mainly to point out that realistic results on the binding energy could be obtained if one were able to improve the estimates of various constants. In the present situation we do not have any bounds on $E_0(N)$ (other than the simple one $E_0(N) > mc^2N$), but we expect something like $E_0(N)/N \sim mc^2 + (\text{const.}) \hbar c \Lambda$ for not too small Λ .

We set $\mathcal{T}'' = P^+ \sum_{i=1}^N |D_i(A)| P^+$. With $P^+ D(A) P^+$ in place of $|\sigma_j \cdot (p_j + \sqrt{\alpha} A(x_j))|$, (37) is valid for $P^+ V_c P^+$ —in the same way that the inequalities of Theorems 2.1 and 2.2 of ref. 3 are valid with the same constants. We find that

$$\begin{aligned} \frac{1}{\hbar c} H_N^{\text{rel}} &\geq -2\pi\Omega\alpha\mathcal{T}'' - 2\pi\Omega\xi\alpha^2 H_f - 2\pi\Omega\alpha\Gamma N + \mathcal{T}'' + H_f \\ &= (1 - 2\pi\Omega\alpha) \mathcal{T}'' + (1 - 2\pi\Omega\xi\alpha^2) H_f - 2\pi\Omega\alpha\Gamma N \\ &\geq (1 - 2\pi\Omega\xi\alpha^2)(\mathcal{T}'' + H_f) - 2\pi\Omega\alpha\Gamma N \end{aligned} \quad (45)$$

since $\xi\alpha < 1$. Therefore,

$$\frac{\Delta E(N)}{N} \leq 2\pi\Omega\alpha \frac{E_0(N)}{N} + 2\pi\Omega\alpha\Gamma\hbar c. \quad (46)$$

We note that—apart from the unnaturally large constant—the binding energy appears to be bounded by α times the self-energy. We also note that (45) and (46) can be improved a little by using the free parameter $0 \leq \varepsilon < 1$ that appears in Section 4 of ref. 3; we have taken $\varepsilon = 0$ here, as we did in Section 3.

ACKNOWLEDGMENTS

E.H.L. was partially supported by U.S. National Science Foundation Grant PHY 98-20650-A02 and M.L. was partially supported by U.S. National Science Foundation Grant DMS 00-70589.

REFERENCES

1. L. Bugliaro, J. Fröhlich, and G. M. Graf, Stability of quantum electrodynamics with nonrelativistic matter, *Phys. Rev. Lett.* **77**:3494–3497 (1996).
2. C. Fefferman, J. Fröhlich, and G. M. Graf, Stability of ultraviolet cutoff quantum electrodynamics with non-relativistic matter, *Comm. Math. Phys.* **190**:309–330 (1997).

3. E. H. Lieb and M. Loss, Stability of a model of relativistic quantum electrodynamics, *Comm. Math. Phys.* **228**:561–588 (2002).
4. E. H. Lieb, M. Loss, and J. P. Solovej, Stability of matter in magnetic fields, *Phys. Rev. Lett.* **75**:985–989 (1995).
5. E. H. Lieb, H. Siedentop, and J. P. Solovej, Stability and instability of relativistic electrons in magnetic fields, *J. Stat. Phys.* **89**:37–59 (1997).
6. E. H. Lieb and H.-T. Yau, The stability and instability of relativistic matter, *Comm. Math. Phys.* **118**:177–213 (1988). See also: many-body stability implies a bound on the fine structure constant, *Phys. Rev. Lett.* **61**:1695–1697 (1988).
7. M. Dresden, *H. A. Kramers, Between Tradition and Revolution* (Springer-Verlag, 1987).
8. J. Avron, I. Herbst, and B. Simon, Schrödinger operators with magnetic fields. I. General interactions, *Duke Math. J.* **45**:847–883 (1978).
9. J. M. Combes, R. Schrader, and R. Seiler, Classical bounds and limits for energy distributions of Hamilton operators in electromagnetic fields, *Ann. Phys.* **111**:1–18 (1978).
10. E. H. Lieb and W. E. Thirring, Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities, in *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*, E. H. Lieb, B. Simon, and A. S. Wightman, eds. (Princeton University Press, Princeton, 1976).
11. E. H. Lieb, M. Loss, and H. Siedentop, Stability of relativistic matter via Thomas–Fermi theory, *Helv. Phys. Acta* **69**:974–984 (1996). (A few minor errors have been corrected in the version in the book: W. Thirring, ed. *The Stability of Matter: From Atoms to Stars, Selecta of E. H. Lieb*, 3rd ed. (Springer, 2001)).
12. E. H. Lieb and M. Loss, Self-energy of electrons in non-perturbative QED, in *Differential Equations and Mathematical Physics, University of Alabama, Birmingham, 1999*, R. Weikard and G. Weinstein, eds. (Amer. Math. Soc./Internat. Press, 2000), pp. 255–269. arXiv math-ph/9908020, mp_arc 99-305. (A few errors have been corrected in the version in the book: W. Thirring, ed., *The Stability of Matter: From Atoms to Stars, Selecta of E. H. Lieb*, 3rd ed. (Springer, 2001)).
13. E. H. Lieb and M. Loss, The ultraviolet problem in non-relativistic QED, in preparation.

Stability of a Model of Relativistic Quantum Electrodynamics*

Elliott H. Lieb^{1, **}, Michael Loss^{2, ***}

¹ Departments of Physics and Mathematics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, NJ 08544, USA

² School of Mathematics, Georgia Tech, Atlanta, GA 30332, USA

Received: 8 September 2001 / Accepted: 18 March 2002

Abstract: The relativistic “no pair” model of quantum electrodynamics uses the Dirac operator, $D(A)$ for the electron dynamics together with the usual self-energy of the quantized ultraviolet cutoff electromagnetic field A – in the Coulomb gauge. There are no positrons because the electron wave functions are constrained to lie in the positive spectral subspace of some Dirac operator, D , but the model is defined for any number, N , of electrons, and hence describes a true many-body system. In addition to the electrons there are a number, K , of fixed nuclei with charges $\leq Z$. If the fields are not quantized but are classical, it was shown earlier that such a model is always unstable (the ground state energy $E = -\infty$) if one uses the customary $D(0)$ to define the electron space, but is stable ($E > -\text{const.}(N + K)$) if one uses $D(A)$ itself (provided the fine structure constant α and Z are not too large). This result is extended to quantized fields here, and stability is proved for $\alpha = 1/137$ and $Z \leq 42$. This formulation of QED is somewhat unusual because it means that the electron Hilbert space is inextricably linked to the photon Fock space. But such a linkage appears to better describe the real world of photons and electrons.

1. Introduction

The theory of the ground state of matter interacting with Coulomb forces and with the magnetic field is not yet in a completely satisfactory state. Open problems remain, such as the inclusion of relativistic mechanics into the many-body formalism and the inclusion of the self-energy effects of the radiation field, especially the quantized radiation field.

One of the fundamental attributes of quantum mechanics is the existence of a Hamiltonian with a lowest, or ground state energy, and not merely the existence of a critical

* © 2001 by the authors. This paper may be reproduced, in its entirety, for non-commercial purposes.

** Work partially supported by U.S. National Science Foundation grant PHY 98-20650-A02.

*** Work partially supported by U.S. National Science Foundation grant DMS 00-70589.

point of a Lagrangian. The “stability” problem, which concerns us here, is to show that the ground state energy is bounded below by a constant times the total number of particles, $N + K$, where N is the number of electrons and K is the number of nuclei – whose locations, in this model, are fixed, but chosen to minimize the energy. We do not discuss the existence of a normalizable ground state eigenfunction, as in [8], but only the lower boundedness of the Hamiltonian.

This problem has been resolved successfully in various models such as the usual nonrelativistic Schrödinger Hamiltonian with only electrostatic interactions. Further developments include extensions to relativistic kinetic energy $\sqrt{p^2 + m^2} - m$ in place of the nonrelativistic $p^2/2m$, and extensions to matter interacting with classical magnetic fields (including a spin-field interaction B), stabilized by the classical field energy

$$\frac{1}{8\pi} \int B(x)^2 dx, \quad (1)$$

and then the quantization of the B field. Many people participated in this development and we refer the reader to [17] and the references therein for an account up to 1997.

In this paper we take a further step by addressing the problem of *relativistic* matter, using the *Dirac operator* (without pair production, i.e., the “no-pair” model) interacting with the *quantized* radiation field having an ultraviolet cutoff Λ . In [17] the corresponding problem was solved with a classical radiation field, in which the field energy is given by (1), and we shall use some of the ideas of that paper here. The idea for such a model goes back to [3] and [23]. With a classical B field no ultraviolet cutoff is needed, but it is needed with a quantized field, for otherwise the field energy diverges.

Because of the ultraviolet cutoff our model, which in other respects is relativistic, is not truly relativistic at energies of the order of the cutoff. We have not attempted to renormalize the theory but, if this can be done consistently, the resulting theory will be relativistic at all energies.

In [4] the problem of *nonrelativistic* electrons (with spin) interacting with the quantized ultraviolet cutoff field was solved by using results in [16] but using only the part of the field energy within a distance $1/\Lambda$ of the fixed nuclei. The constants and exponents in [4] were improved in [7]; in particular, the Hamiltonian is bounded below by $-\Lambda K$. The relation of the classical field energy to the quantized field energy involves a commutator that, when integrated over the whole space $\mathbb{R}^3$ yields an infinite constant, even with an ultraviolet cutoff. This is the reason for considering only a local field energy, since only a local field energy yields a finite commutator, and we do the same here.

In Sect. 2 our model is defined and the main Theorem 2.1 is stated. With the fine structure constant $\alpha = 1/137$, stability holds for $Z \leq 42$. The main idea of the “no-pair” model is that there are no positrons, and electronic wave functions are allowed to lie only in the positive spectral subspace of some Dirac operator D . While the Dirac operator $D(A)$, which is contained in the Hamiltonian and which defines the electron dynamics, always contains the magnetic vector potential $A(x)$, the operator D that defines an electron could be $D(0)$, the free Dirac operator. Indeed, this is the conventional choice, but it is not gauge invariant and *always* leads to instability as first shown in [17] for classical fields and here for quantized fields.

The question of instability is complicated. There are two kinds (first and second) and two cases to consider (with and without Coulomb potentials). Instability of the first kind means that the ground state energy (bottom of the spectrum of the Hamiltonian) is $-\infty$. Instability of the second kind means that the energy is finite but is not bounded below

by a constant times $N + K$. The occurrence of these instabilities may or may not depend on α and Z and whether or not a cutoff Λ is present.

The physical nature of the instability, if it occurs, is different in the two cases. When it occurs in the absence of Coulomb potentials (meaning that the αV_c term in (11) is omitted) it is due to the $\sqrt{\alpha} A(x)$ term in $D(A)$ blowing up. When it occurs because of the Coulomb potentials being present it is due to an electron falling into the Coulomb singularity of the nucleus. The various possibilities, all proved in this paper, are summarized in detail in the following two tables and discussed in detail in Appendix E. For the proofs of the instabilities listed here, we rely heavily on [17] and [9].

Table 1. Electrons defined by projection onto the positive subspace of $D(0)$, the free Dirac operator

	Classical or quantized field without cutoff Λ $\alpha > 0$ but arbitrarily small	Classical or quantized field with cutoff Λ $\alpha > 0$ but arbitrarily small
Without Coulomb potential αV_c	Instability of the first kind	Instability of the second kind
With Coulomb potential αV_c	Instability of the first kind	Instability of the second kind

Table 2. Electrons defined by projection onto the positive subspace of $D(A)$, the Dirac operator with field

	Classical field with or without cutoff Λ or quantized field with cutoff Λ
Without Coulomb potential αV_c	The Hamiltonian is positive
With Coulomb potential αV_c	Instability of the first kind when either α or $Z\alpha$ is too large Stability of the second kind when both α and $Z\alpha$ are small enough

The main point of this paper is the proof of the bottom row of the second table in the quantized case; the classical case was done in [17].

There are several ways in which one could hope to go further. One is that one should really prove stability for the *binding* energy, i.e., one should compute the energy difference between that of free particles and that of the interacting system. In a theory with quantized fields the self-energy, i.e., the energy of a free electron, is unknown and quite large. As we show in [13] and [14] the self-energy of a nonrelativistic particle with spin is bounded below by $+\Lambda$, and probably even $+\Lambda^{3/2}$. Moreover, for N fermions (but *not* for bosons) this energy is proportional to $C'N\Lambda$ with $C' > 0$.

Another very important problem to consider is renormalization; our mass m is the unrenormalized one. An answer to this problem also has to address the question of the meaning of mass in an ultraviolet cut-off model, since several definitions are possible. Is it the coefficient of β in an effective Dirac operator that gives the renormalized dynamics, or is it the ground state energy of a “bound” electron?

The results in this paper will be used in another paper of ours [15] to give upper bounds to the hydrogen atom binding energy (and hence to the mass renormalization

using the first definition) in this relativistic no-pair model and in some non-relativistic models with quantized fields.

Finally, let us note that the inclusion of positrons into the model cannot change the fact that defining an electron by means of $D(0)$ will still cause the instabilities listed in the tables above. The reason is simply that the existence of positrons does not prevent one from considering states consisting purely of electrons, and these alone can cause the listed instabilities.

The use of $D(A)$ instead of $D(0)$ to define the electron requires a significant change in the Hilbert space structure of QED. It is no longer possible to separate the Hilbert space for the electron coordinates from the Hilbert space (Fock space) of the photons. The two are now linked in a manner that we describe in the next section.

2. Basic Definitions

We consider N relativistic electrons in the field of K nuclei, fixed at the positions $R_1, \dots, R_K \in \mathbb{R}^3$. (In the real world the fixed nuclei approximation is a good one since the masses of the nuclei are so large compared to the electron's mass.) We assume that their atomic numbers $Z_1, \dots, Z_K$ are all less than some fixed number $Z > 0$. Since the energy is a concave function of each Z_j separately, it suffices, for finding a lower bound, either to put $Z_j = 0$, i.e., to remove the j^{th} nucleus, or to put $Z_j = Z$ (see [5]). Thus, without loss of generality, we may assume that all the nuclear charges are equal to Z .

We use units in which $\hbar = 1$ and $c = 1$. $\alpha = e^2/\hbar c$ is the dimensionless “fine structure constant” ($=1/137$ in nature). The electric charge of the electron in these units is $e = \sqrt{\alpha}$.

We use the Coulomb, or radiation gauge so that the Coulomb potential is a function only of the coordinates of the N electrons, $x_1, x_2, \dots, x_N$ and equals αV_c , where

$$V_c = -Z \sum_{i=1}^N \sum_{k=1}^K \frac{1}{|x_i - R_k|} + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|} + Z^2 \sum_{1 \leq k < l \leq K} \frac{1}{|R_k - R_l|}. \tag{2}$$

In this gauge, it is the vector potential that is quantized. A careful discussion of the field and its quantization is given in Appendix A. The (ultraviolet cutoff) magnetic vector potential is defined by

$$A(x) = \frac{1}{2\pi} \sum_{\lambda=1}^2 \int_{|k| \leq \Lambda} \frac{\varepsilon_\lambda(k)}{\sqrt{\omega(k)}} \left(a_\lambda(k) e^{ik \cdot x} + a_\lambda^*(k) e^{-ik \cdot x} \right) dk, \tag{3}$$

where Λ is the ultraviolet cutoff on the wave-numbers $|k|$. The operators a_λ, a_λ^* satisfy the usual commutation relations

$$[a_\lambda(k), a_\nu^*(q)] = \delta(k - q) \delta_{\lambda, \nu}, \quad [a_\lambda(k), a_\nu(q)] = 0, \quad \text{etc.}, \tag{4}$$

and the vectors $\varepsilon_\lambda(k)$ are the two possible orthonormal polarization vectors perpendicular to k and to each other.

Our results hold for all finite Λ . The details of the cutoff in (3) are quite unimportant, except for the requirement that rotation symmetry in k -space is maintained. E.g., a

Gaussian cutoff can be used instead of our sharp cutoff. We avoid unnecessary generalizations. The cutoff resides in the A -field, not in the field energy, H_f , sometimes called $d\Gamma(\omega)$, which is given by

$$H_f = \sum_{\lambda=1,2} \int_{\mathbb{R}^3} \omega(k) a_{\lambda}^*(k) a_{\lambda}(k) dk. \quad (5)$$

The energy of a photon is $\omega(k)$ and the physical value of interest to us, which will be used in the rest of this paper, is

$$\omega(k) = |k|. \quad (6)$$

Again, generalizations are possible, but we omit them.

An important fact for our construction of the physical Hilbert space of our model is that $[A(x), A(y)] = [B(x), B(y)] = [A(x), B(y)] = 0$ for all x, y . Here, B is the magnetic field given by

$$B(x) = \text{curl } A(x) = \frac{i}{2\pi} \sum_{\lambda=1}^2 \int_{|k| \leq \Lambda} \frac{k \wedge \varepsilon_{\lambda}(k)}{\sqrt{\omega(k)}} \left(a_{\lambda}(k) e^{ik \cdot x} - a_{\lambda}^*(k) e^{-ik \cdot x} \right) dk. \quad (7)$$

The kinetic energy of an electron is defined in terms of a Dirac operator with the vector potential $A(x)$ (with x being the electron's coordinate)

$$D(A) := \alpha \cdot (-i\nabla + \sqrt{\alpha} A(x)) + m\beta, \quad (8)$$

with α and β given in terms of the 2×2 Pauli matrices and 2×2 identity I as

$$\alpha = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note that

$$D(A)^2 = \widehat{T}^P(A) + m^2, \quad (9)$$

where $\widehat{T}^P(A) = \begin{pmatrix} T^P(A) & 0 \\ 0 & T^P(A) \end{pmatrix}$ and $T^P(A)$ is the Pauli operator on $L^2(\mathbb{R}^3; \mathbb{C}^2)$,

$$T^P(A) = [\sigma \cdot (p + \sqrt{\alpha} A(x))]^2 = (p + \sqrt{\alpha} A(x))^2 + \sqrt{\alpha} \sigma \cdot B(x). \quad (10)$$

As a step towards defining a physical Hamiltonian for our system of N electrons and K fixed nuclei, we first define a conventional, but fictitious Hamiltonian

$$H'_N = \sum_{i=1}^N D_i(A) + \alpha V_c + H_f. \quad (11)$$

This H'_N acts on the usual Hilbert space $\mathcal{H}_N = \bigotimes^N L^2(\mathbb{R}^3; \mathbb{C}^4) \otimes \mathcal{F}$, where $\mathcal{F}$ is the Fock space for the A -field. A vector in $\mathcal{H}_N$ can be written

$$\Psi = \bigoplus_{j=0}^{\infty} \Phi_j(x_1, \dots, x_N; \tau_1, \dots, \tau_N; k_1, \dots, k_j; \lambda_1, \dots, \lambda_j). \quad (12)$$

Here, the λ_i take the two values 1, 2 and the τ_i take the four values 1, 2, 3, 4. Each Φ_j is symmetric in the pairs of variables k_i, λ_i and it is square integrable in x, k . The sum of these integrals (summed over λ 's, τ 's, and j) is finite. The operators $a_\lambda(k)$ and their adjoints act, as usual, by

$$a_\lambda(k)\Psi = \bigoplus_{j=0}^{\infty} \sqrt{j+1} \Phi_{j+1}(x_1, \dots, x_N; \tau_1, \dots, \tau_N; k_1, \dots, k_j, k; \lambda_1, \dots, \lambda_j, \lambda). \quad (13)$$

As mentioned before, the physical Hilbert space is constructed using the positive spectral projections of the Dirac operators $D_i(A)$. By Lemma C.1 the N Dirac operators commute in the strong sense that their spectral projections commute with each other. Thus, the Hilbert space $\mathcal{H}_N$ can be divided into 2^N subspaces according to the positive and negative spectral subspaces of each $D_i(A)$. (Note that as long as $m > 0$ there is no zero spectral subspace.) We denote by P^+ the orthogonal projection onto the positive spectral subspace for *all* the Dirac operators.

The space $P^+\mathcal{H}_N$ is invariant (up to unitary equivalence) by the natural action of the permutation group S_N consisting of permutations of the electron labels. In accordance with the Pauli principle we choose the antisymmetric component of $P^+\mathcal{H}_N$, as the physical Hilbert space. Thus, our physical Hilbert space is given as

$$\mathcal{H}_N^{\text{phys}} = \mathcal{A} P^+ \mathcal{H}_N, \quad (14)$$

where $\mathcal{A}$ is the projector onto the antisymmetric component.

Formally, i.e., without attention to domain questions, our physical Hamiltonian on $\mathcal{H}_N^{\text{phys}}$ is defined to be

$$H_N^{\text{phys}} = P^+ H'_N P^+. \quad (15)$$

Since we are interested in this operator as a quadratic form, it suffices to specify a domain $\mathcal{Q}_N$ which is dense in $\mathcal{H}_N^{\text{phys}}$ and on which the expectation values of all the operators involved are finite. Since all the operators are symmetric, and since a stability estimate entails that the quadratic form is bounded below, its closure exists and defines a selfadjoint operator H_N^{phys} . Such a domain $\mathcal{Q}_N$ is constructed in Appendix D. Note that by definition $\mathcal{Q}_N$ consists of antisymmetric elements.

We note that each of the N Dirac operators commute with P^+ . For $\psi \in \mathcal{Q}_N$ we have $D_i(A)\psi = P^+ D_i(A)\psi$. For the other two terms in (11) the role of the projector is not so trivial and that is why we have to write $P^+ H'_N P^+$.

This model has its origins in the work of Brown and Ravenhall [3] and Sucher [23]. The immediate antecedent is [17].

Let us note five things:

(i) It is not entirely easy to think about $\mathcal{H}_N^{\text{phys}}$ because the electronic L^2 -spaces and the Fock space are now linked together. In our choice of positive energy states, the electrons have their own photon cloud. We chose to apply the projector P^+ first and then antisymmetrize. As explained in Appendix D, we can, of course, do it the other way around and obtain the same Hilbert space, since P^+ commutes with permutations. We also show in Appendix D that $\mathcal{H}_N^{\text{phys}}$ is not trivial; in fact it is infinite dimensional.

(ii) Usually, in quantum electrodynamics, one defines $\mathcal{H}_N^{\text{phys}}$ by means of the positive spectral subspace of the *free* Dirac operator $D(0) = -i\alpha \cdot \nabla + m\beta$, instead of $D(A)$.

This is easier to think about but, as demonstrated in [17] with a classical A field instead of a quantized field, the choice of $D(0)$ always leads to instability, as listed in the tables in Sect. 1 and discussed in detail in Appendix E.

(iii) Because of the restriction to the positive spectral subspace of $D(A)$, the Dirac operator is never negative. The only negative terms in H_N^{phys} come from the Coulomb potential. It should also be noted that the choice of the free Dirac operator to define an electronic wave function is not a gauge covariant notion. The $D(A)$ choice is gauge covariant.

(iv) $\mathcal{H}_N^{\text{phys}}$ depends on α and m .

(v) While energy, being one component of a four-vector, is not a relativistically invariant quantity, it is true, nevertheless, that positive and negative energies of $D(A)$ are relativistic concepts since they are invariant under Lorentz transformations that do not change the direction of time. We thank J-M. Graf for this remark and we thank J. Yngvason for noting that for this to be true it is essential that the joint spectrum of energy and momentum of $D(A)$ lies in the light cone. We have not proved this, but it is plausibly true.

Our main result, to be proved in Sect. 4, is

Theorem 2.1 (Relativistic quantum electrodynamic stability). *Assume that Z and α are such that there is a solution κ and $\varepsilon \geq 0$ to the three inequalities (52), (53) and (54). Then H_N^{phys} in (15) is bounded below by*

$$H_N^{\text{phys}} \geq +\sqrt{\varepsilon} m N - \frac{18\Lambda}{\pi} K C_2^3, \quad (16)$$

where

$$C_2^4 = \frac{N}{K} \frac{6\sqrt{1-\varepsilon} + (\alpha/2)(\sqrt{2Z} + 2.3)^2}{27/2\pi}. \quad (17)$$

In particular, $Z \leq 42$ is allowed when $\alpha = 1/137$.

Actually, our proof of Theorem 2.1 utilizes the absolute value of the Dirac operator $|D(A)|$ on the Hilbert space $\mathcal{H}_N$. If we recall the connection between $D(A)$ and the Pauli operator in (9) and (10), we can prove the following theorem as a byproduct of our proof of Theorem 2.1. Note $\mathbb{C}^2$ in place of $\mathbb{C}^4$ here.

Theorem 2.2 (Stability with the Pauli operator). *Let $H_N'' = \sum_{i=1}^N \sqrt{T^P(A) + m^2} + \alpha V_c + H_f$ be a Hamiltonian on the space $\mathcal{A} \otimes^N L^2(\mathbb{R}^3; \mathbb{C}^2) \otimes \mathcal{F}$. Then stability of the second kind holds under the conditions stated in Theorem 2.1, and with the same lower bound (16).*

3. Bounding the Coulomb Potential by a Localized Relativistic Kinetic Energy

The following Theorem 3.1 contains the main technical estimate needed in this paper, but it is independently interesting. It deals with a model of relativistic electrons interacting with quantized fields, but without the spin-field interaction and without the field energy. While this model is different from the no-pair Hamiltonian (15), some of its properties will be useful later. We consider two such Hamiltonians: A usual one

$$\kappa \sum_{i=1}^N |p_i + \sqrt{\alpha} A(x_i)| + V_c, \quad (18)$$

(with $\kappa > 0$) and a related one with a localized kinetic energy described below in (24), (25). In this section $A(x)$ is some given *classical* field, not necessarily divergence free. There is no α in front of V in (18). The Hilbert space is $\mathcal{A} \otimes^N L^2(\mathbb{R}^3; \mathbb{C}^q)$ for fermions with q “spin states”.

With the K nuclei positioned at distinct points $R_j \in \mathbb{R}^3$, for $j = 1, \dots, K$, we define the corresponding Voronoi cells by

$$\Gamma_j = \{x \in \mathbb{R}^3 : |x - R_j| < |x - R_i|, i = 1, \dots, K, i \neq j\}. \tag{19}$$

These Voronoi cells are open convex sets. We choose some $L > 0$ and define the balls $\mathcal{B}_j \subset \mathbb{R}^3$ by

$$\mathcal{B}_j = \{x : |x - R_j| \leq 3L\}, \tag{20}$$

and denote by $\mathcal{B}$ the union of these K balls and by $\chi_{\mathcal{B}}$ the characteristic function of $\mathcal{B}$. Similarly, we define smaller balls, $\mathcal{S}_j = \{x : |x - R_j| \leq 2L\}$, and define $\chi_{\mathcal{S}}$ to be the characteristic function of the union of these K smaller balls. Choose some function $g \in W^{1,1}(\mathbb{R}^3)$ with support in $\{x : |x| \leq 1\}$, with $g \geq 0$ and with $\int g = 1$. Define $g_L(x) = L^{-3}g(x/L)$. Clearly $\int g_L = 1$ and g_L has support in $\{x : |x| \leq L\}$. With $*$ denoting convolution, set

$$\phi_1(x) = g_L * \chi_{\mathcal{S}}(x). \tag{21}$$

This function ϕ_1 is nonnegative and everywhere bounded by 1. We also define $\phi_2 = 1 - \phi_1$ and set

$$F = \phi_1 / \sqrt{\phi_1^2 + \phi_2^2}, \quad \text{and} \quad G = \phi_2 / \sqrt{\phi_1^2 + \phi_2^2}. \tag{22}$$

Note that $\phi_1^2 + \phi_2^2 \geq 1/2$ and $F(x) = 1$ if $|x - R_j| < L$ for some j . Note also that F and ϕ_1 are supported in $\mathcal{B}$, i.e., $\chi_{\mathcal{B}}\phi_1 = \phi_1$ and $\chi_{\mathcal{B}}F = F$.

We find that

$$|\nabla F|^2 + |\nabla G|^2 \leq 4|\nabla \phi_1|^2 \leq \frac{4}{L^2} \left(\int_{\mathbb{R}^3} |\nabla g(x)| dx \right)^2, \tag{23}$$

and hence $|\nabla F|, |\nabla G| \leq 2|\nabla \phi_1|$.

The function g that minimizes the integral in (23) is $g(x) = 3/4\pi$ for $|x| \leq 1$ and zero otherwise. (Although this g is not in $W^{1,1}(\mathbb{R}^3)$, it is a limit of $W^{1,1}(\mathbb{R}^3)$ functions.) Then the integral equals 3 and $|\nabla F|^2 + |\nabla G|^2 \leq 36L^{-2}$.

The localized kinetic energy operator $Q(A)$ is given by

$$Q(A) = F(x) |p + \sqrt{\alpha} A(x)| F(x) = F(x) \sqrt{(p + \sqrt{\alpha} A(x))^2} F(x). \tag{24}$$

This operator is well defined as a quadratic form since the function F is smooth, and hence defines a self adjoint operator via Friedrich’s extension.

The related relativistic Hamiltonian, with localized kinetic energy, is now defined by

$$H_N^{\text{loc}} := \kappa \sum_{i=1}^N Q_i(A) + V_c, \tag{25}$$

and has the following bound which, it is to be noted, does not depend on the details of $g(x)$.

Theorem 3.1 (Bound on Coulomb energy). *For any vector field $A(x)$ and for N fermions with q spin states,*

$$\kappa \sum_{i=1}^N Q_i(A) + V_c \geq -\frac{N}{2L} \max \left\{ (\sqrt{2Z} + 1)^2, 2Z + \frac{110}{21} \right\} \geq -\frac{N}{2L} (\sqrt{2Z} + 2.3)^2, \quad (26)$$

provided $\kappa \geq \max\{q/0.031, \pi Z\}$.

Proof. It was proved in [19] (Eqs. (2.4)–(2.6) with $\lambda = 10/11$) that the Coulomb potential V_c is bounded below by a *single-particle* potential plus a constant, namely, for $x_i, R_j \in \mathbb{R}^3$,

$$V_c \geq -\sum_{i=1}^N W(x_i) + \frac{Z^2}{8} \sum_{j=1}^K \frac{1}{D_j}, \quad (27)$$

where $2D_j = \min_{i \neq j} \{|x_i - x_j|\}$ and, for $x \in \Gamma_j$,

$$\begin{aligned} W(x) &= \frac{(\sqrt{Z} + 1/\sqrt{2})^2}{|x - R_j|} \text{ for } |x - R_j| \geq \frac{10D_j}{11} \\ &= \frac{Z}{|x - R_j|} + \frac{121}{42D_j} \text{ for } |x - R_j| < \frac{10D_j}{11}. \end{aligned} \quad (28)$$

This estimate reduces our problem to finding a lower bound to

$$\kappa \sum_{i=1}^N Q_i(A) - \sum_{i=1}^N F(x_i)^2 W(x_i) - \sum_{i=1}^N (1 - F(x_i)^2) W(x_i) + \frac{Z^2}{8} \sum_{j=1}^K \frac{1}{D_j}. \quad (29)$$

Since $F(x) = 1$ if $|x - R_j| < L$ for some j , the third term in (29) is bounded below by

$$-\frac{N}{2L} \max\{(\sqrt{2Z} + 1)^2, 2Z + \frac{110}{21}\}. \quad (30)$$

Estimating the first and second terms using the Pauli exclusion principle amounts to filling the lowest possible energy levels with q electrons each, and this energy is bounded below by q times the sum of the negative eigenvalues of the operator

$$F(x)(|p + \sqrt{\alpha}A(x)| - W(x))F(x). \quad (31)$$

According to the generalized min-max principle [12] Corollary 12.2, and the fact that $\|F\psi\| \leq \|\psi\|$, this is bounded below by q times the sum of the negative eigenvalues of the operator $|p + \sqrt{\alpha}A(x)| - W(x)$. However, Theorem 1 of [19], shows that this sum is not less than $(-Z^2/8) \sum_{j=1}^K 1/D_j$ under the stated condition on κ .

(Notes: We refer here to Theorem 1 of [19] because, as noted in [17], the proof of that theorem holds for $|p + \sqrt{\alpha}A(x)|$ in place of $|p|$. While Theorem 1 of [19] is stated in terms of V_c , the proof in [19] actually replaces V_c by its lower bound (27).) $\square$

4. Proof of Theorem 2.1

We employ a strategy similar to that in [17].

As a first step we use Theorem 3.1 with a suitable choice of L to control the Coulomb potential.

The operators appearing in Theorem 3.1 do not involve spin, but the number of spin states, q , is important for determining the relevant value of κ . The correct choice is $q = 2$, not $q = 4$, as explained in [17], p. 42 and Appendix B. The point is the following. The one-body density matrix $\Gamma(x, \sigma; x', \sigma')$ coming from an antisymmetric N particle wave function Ψ defines a reduced one body density matrix

$$\gamma(x, x') = \sum_{\sigma=1}^4 \Gamma(x, \sigma; x', \sigma). \tag{32}$$

This reduced density matrix, in general, satisfies $0 \leq \text{Tr} \gamma \equiv \int \gamma(x, x) dx \leq 4$. If, however, Ψ is in the range of P^+ , then $0 \leq \text{Tr} \gamma \leq 2$, as shown in [17]. In the proof of Theorem 3.1, the only relevant information about Ψ enters via the reduced single particle matrix γ . Thus, we require only $\kappa \geq \max\{64.5, \pi Z\}$.

In the definition of F we set $L = C_2/\Lambda$, where $C_2 > 0$ is some constant to be conveniently chosen later. We then have (recalling (9), (10), and $P_i^+ D_i(A) P_i^+ = P_i^+ |D_i(A)| P_i^+$)

$$\begin{aligned} P^+ \left[\sum_{i=1}^N D_i(A) + \alpha V_c \right] P^+ &\geq P^+ \sum_{i=1}^N \left[\sqrt{\widehat{T}_i^P(A) + m^2} - \kappa \alpha Q_i(A) \right] P^+ \\ &\quad - \alpha N \frac{\Lambda}{2C_2} (\sqrt{2Z} + 2.3)^2 P^+. \end{aligned} \tag{33}$$

(Here, $Q(A)$ really denotes the 4×4 operator $Q(A) \otimes I_4$, where I_4 is the identity in spin-space.) Consider the operator

$$H_2 := P^+ \sum_{i=1}^N \left[\sqrt{\widehat{T}_i^P(A) + m^2} - \delta m - \kappa \alpha Q_i(A) + C_3 \Lambda \right] P^+, \tag{34}$$

where the numbers $0 \leq \delta \leq 1$ and $C_3 > 0$ will be chosen later.

If we denote by Ξ^+ the projection onto the positive spectral subspace of $D(A)$ acting on $L^2(\mathbb{R}^3; \mathbb{C}^4) \otimes \mathcal{F}$, then H_2 is bounded below by

$$\text{Tr}_4[\Xi^+ S \Xi^+]_-, \tag{35}$$

where Tr_n with $n = 1, 2, 4$ denotes the trace on $L^2(\mathbb{R}^3; \mathbb{C}^n)$. The operator S is

$$S := \sqrt{\widehat{T}^P(A) + m^2} - \delta m - \kappa \alpha Q_i(A) + C_3 \Lambda. \tag{36}$$

It has the form

$$S = \begin{pmatrix} Y & 0 \\ 0 & Y \end{pmatrix}. \tag{37}$$

Here, the entry Y is a 2×2 matrix valued operator and $[X]_-$ denotes the negative part of a self-adjoint operator X (and which is nonnegative by definition). The projection Ξ^+ is

not explicitly given, but observing, as in [17], that the projection Ξ^- onto the *negative* energy states is related to Ξ^+ by

$$\Xi^- = U^{-1} \Xi^+ U = -U \Xi^+ U, \quad (38)$$

where U is the matrix

$$U = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad (39)$$

we see that the operators $\Xi^+ S \Xi^+$ and $\Xi^- S \Xi^-$ have the same spectrum. Thus,

$$\mathrm{Tr}_4[\Xi^+ S \Xi^+]_- \leq \frac{1}{2} \mathrm{Tr}_4[S]_- = \mathrm{Tr}_2 \left[\sqrt{T^P(A) + m^2} - \delta m - \kappa \alpha Q(A) + C_3 \Lambda \right]_- . \quad (40)$$

Therefore, the infimum of the spectrum of H_2 over states that satisfy the Pauli exclusion principle (with 4 spin states) is bounded below by

$$-\mathrm{Tr}_2 \left[\sqrt{T^P(A) + m^2} - \delta m - \kappa \alpha Q(A) + C_3 \Lambda \right]_- . \quad (41)$$

The BKS inequality [2] (see also [17]) states that for positive operators A and B , $\mathrm{Tr}_2[A - B]_- \leq \mathrm{Tr}_2[A^2 - B^2]_-^{1/2}$. Note that $\sqrt{T^P(A) + m^2} - \delta m \geq 0$ and, therefore,

$$H_2 \geq -\mathrm{Tr}_2 \left[\left(\sqrt{T^P(A) + m^2} - \delta m + C_3 \Lambda \right)^2 - \kappa^2 \alpha^2 Q(A)^2 \right]_-^{1/2}, \quad (42)$$

which is greater than

$$-\mathrm{Tr}_2 \left[\left(\sqrt{T^P(A) + m^2} - \delta m \right)^2 + C_3^2 \Lambda^2 - \kappa^2 \alpha^2 Q(A)^2 \right]_-^{1/2}. \quad (43)$$

(Here, and in the following, we use the fact that $\mathrm{Tr}[X]_-^{1/2}$ is monotone decreasing in X .)

Next, we expand $(\dots)^2$ in (43) and use the arithmetic-geometric mean inequality to bound (43) from below by

$$-\mathrm{Tr}_2 \left[(T^P(A) + m^2)(1 - \varepsilon) + (1 - 1/\varepsilon)m^2\delta^2 + C_3^2 \Lambda^2 - \kappa^2 \alpha^2 Q(A)^2 \right]_-^{1/2}. \quad (44)$$

We choose δ so that the mass disappears, i.e., $\delta^2 = \varepsilon$.

The next step is to localize the Pauli term $T^P(A)$. A standard calculation shows that (with F, G as in Sect. 3)

$$\begin{aligned} T^P(A) &= F T^P(A) F + G T^P(A) G - |\nabla F|^2 - |\nabla G|^2 \\ &\geq F T^P(A) F - |\nabla F|^2 - |\nabla G|^2. \end{aligned} \quad (45)$$

We insert the right side of (45) into (44) and, recalling (23), choose C_3 to eliminate the Λ^2 term, i.e.,

$$C_3 = \frac{2\sqrt{1-\varepsilon}}{C_2} \left(\int_{\mathbb{R}^3} |\nabla g(x)| dx \right) = \frac{6\sqrt{1-\varepsilon}}{C_2}. \quad (46)$$

Thus, using the fact that $Q(A)^2 = F|p + \sqrt{\alpha}A(x)|F^2|p + \sqrt{\alpha}A(x)|F \leq F(p + \sqrt{\alpha}A(x))^2F$, which follows from $F^2 \leq 1$, we obtain the bound

$$\begin{aligned} H_2 &\geq -\mathrm{Tr}_2 \left[(1 - \varepsilon)FT^P(A)F - \kappa^2\alpha^2Q(A)^2 \right]_-^{1/2} \\ &\geq -\mathrm{Tr}_2 \left[F \left((1 - \varepsilon - \kappa^2\alpha^2)(p + \sqrt{\alpha}A(x))^2 + (1 - \varepsilon)\sqrt{\alpha}\chi_B\sigma \cdot B(x) \right) F \right]_-^{1/2}. \end{aligned} \tag{47}$$

We have used the fact that $\chi_B F = F$.

Since $FXF \geq -F[X]_-F$ for any X , the eigenvalues of FXF are bounded below by the eigenvalues of $-F[X]_-F$, and hence we have that $\mathrm{Tr} [FXF]_-^{1/2} \leq \mathrm{Tr} [F X_- F]^{1/2}$, and hence

$$\begin{aligned} \mathrm{Tr}_2 \left[F \left\{ (1 - \varepsilon - \kappa^2\alpha^2)(p + \sqrt{\alpha}A(x))^2 + (1 - \varepsilon)\sqrt{\alpha}\chi_B\sigma \cdot B(x) \right\} F \right]_-^{1/2} \\ \leq \mathrm{Tr}_2 \left\{ F \left[(1 - \varepsilon - \kappa^2\alpha^2)(p + \sqrt{\alpha}A(x))^2 + (1 - \varepsilon)\sqrt{\alpha}\chi_B\sigma \cdot B(x) \right]_- F \right\}^{1/2}. \end{aligned}$$

The expression $[\]_-$ between the two F 's is, by definition, a positive-semidefinite self-adjoint operator and we denote it by Y . Now

$$\mathrm{Tr}_2(FYF)^{1/2} = \mathrm{Tr}_2(FY^{1/2}Y^{1/2}F)^{1/2} = \mathrm{Tr}_2(Y^{1/2}FFY^{1/2})^{1/2} \tag{48}$$

since, quite generally, X^*X and XX^* have the same spectrum (up to zero eigenvalues, which are not counted here). Finally, we note that since $F^2 \leq 1$, $Y^{1/2}FFY^{1/2} \leq Y$, and hence $\mathrm{Tr}_2(FYF)^{1/2} = \mathrm{Tr}_2(Y^{1/2}FFY^{1/2})^{1/2} \leq \mathrm{Tr}_2Y^{1/2}$. Thus, it remains to find an upper bound to $[h]_-^{1/2}$, where

$$h = (1 - \varepsilon - \kappa^2\alpha^2)(p + \sqrt{\alpha}A(x))^2 + (1 - \varepsilon)\sqrt{\alpha}\chi_B\sigma \cdot B(x). \tag{49}$$

Denote the negative eigenvalues of h by $-e_1 \leq -e_2 \leq \dots$. One way to bound the eigenvalues from below is to replace $\sigma \cdot B(x)$ by $-|B(x)|$, but then each eigenvalue of $\tilde{h} := (1 - \varepsilon - \kappa^2\alpha^2)(p + \sqrt{\alpha}A(x))^2 - (1 - \varepsilon)\sqrt{\alpha}\chi_B|B(x)|$ on $L^2(\mathbb{R}^3)$ would have to be counted twice (because Tr_2 is over $L^2(\mathbb{R}^3; \mathbb{C}^2)$ and not $L^2(\mathbb{R}^3)$). As shown in [20], however, the intuition that each negative eigenvalue of $\tilde{h}$ should be counted only once is correct in that the Lieb–Thirring inequality [18] applies and we obtain the bound

$$\sum_i \sqrt{e_i} \leq \frac{(1 - \varepsilon)^2 \ell \alpha}{(1 - \varepsilon - \kappa^2\alpha^2)^{3/2}} \int_B B(x)^2 dx, \tag{50}$$

with $\ell = 0.060$ [20].

It is to be emphasized that (50) is an operator inequality. That is, the operator in (34), which is part of H_N^{phys} , satisfies

$$H_2 \geq -\frac{(1 - \varepsilon)^2 \ell \alpha}{(1 - \varepsilon - \kappa^2\alpha^2)^{3/2}} \int_B B(x)^2 dx. \tag{51}$$

The right side of (51) can be controlled by the field energy through inequality (79) – provided $1/8\pi$ is not less than the constant in (50), (51).

4.1. Evaluation of Constants. We are now ready to list the conditions on the constants C_2 and ε that have been introduced and to use these to verify the results of Theorem 2.1,

$$\kappa \geq \max\{64.5, \pi Z\}, \quad (52)$$

$$\text{Conditions :} \quad (\kappa\alpha)^2 < 1 - \varepsilon \leq 1, \quad (53)$$

$$\frac{(1 - \varepsilon)^2 \alpha}{(1 - \varepsilon - \kappa^2 \alpha^2)^{3/2}} \leq \frac{1}{8\pi(0.060)}. \quad (54)$$

The first comes from Theorem 3.1 with $q = 2$. The second is the condition that the kinetic energy term in H_2 is positive. The third is the requirement that the the field energy H_f dominates the sum of the negative eigenvalues in (50).

Assuming these conditions are satisfied the total energy is then bounded below by the sum of the following four terms (recalling (46) and $\delta^2 = \varepsilon$):

$$+\sqrt{\varepsilon} m N, \quad (55)$$

$$-\frac{6\sqrt{1-\varepsilon}}{C_2} \Lambda N, \quad (56)$$

$$\text{Energy Lower Bounds :} \quad -\frac{\alpha \Lambda}{2C_2} \left(\sqrt{2Z} + 2.3\right)^2 N, \quad (57)$$

$$-\frac{\Lambda^4}{8\pi^2} \int_{\mathcal{B}} 1 \geq -\frac{4\pi}{3} \frac{\Lambda}{8\pi^2} (3C_2)^3 K = -\frac{9}{2\pi} \Lambda C_2^3 K. \quad (58)$$

The first comes from the $-\delta m$ term in (34). Similarly, the second comes from the $+C_3 \Lambda$ term in (34). The third term is the last term in (33) which, in turn, comes from Theorem 3.1. The fourth term is the additive constant in (79) with $w(y) = \chi_{\mathcal{B}}(y)$. The volume of $\mathcal{B}$ is bounded by the number of nuclei times the volume of one ball of radius $3L$ around each nucleus.

Obviously we choose

$$C_2^4 = \frac{N}{K} \frac{6\sqrt{1-\varepsilon} + (\alpha/2)(\sqrt{2Z} + 2.3)^2}{27/2\pi}. \quad (59)$$

The sum of the terms (55 – 58) then become our lower bound for the energy

$$\frac{E}{N} \geq +\sqrt{\varepsilon} m - \frac{18\Lambda}{\pi} \frac{K}{N} C_2^3, \quad (60)$$

which satisfies stability of the second kind.

To find the largest possible Z for which stability holds we take $\alpha = 1/137$ and make the choice $\varepsilon = 0$. We then find, from (54), that $\kappa\alpha \leq 0.97$. Setting $\kappa = \pi Z$ we find stability up to $Z = 42$.

The choice $\varepsilon = 0$ makes the energy in (60) negative. Recall that if $Z = 0$ then $E/N = m$. To make contact with physics we would like the energy to be positive, i.e., only a little less than Nm . To fix ideas, let us consider the case $\pi Z \leq 64.5$ and $\alpha = 1/137$. Then $\kappa = 64.5$, $\kappa\alpha = 0.471$ and $(\kappa\alpha)^2 = 0.222$. From (54), we require that (with $x = 1 - \varepsilon \geq 0.222$)

$$x^2 \leq 90.9(x - 0.222)^{3/2}, \quad (61)$$

which means that we can take $1 - \varepsilon = .229$ or $\varepsilon = 0.771$.

Now let us consider the case of hydrogen, $Z = 1$ and $N = K$ (neutrality). From (59) we find that $C_2 = 0.908$. Then (60) becomes

$$\frac{E}{N} = 0.866m - 4.29\Lambda. \quad (62)$$

If Λ is less than one fifth of the electrons' self-energy, the total energy of arbitrarily many hydrogen atoms is positive. This bound could be significantly improved by more careful attention to our various inequalities.

A. Appendix: A Note About Units

The choice of units in electrodynamics is always confusing, especially when interactions with charged particles are involved.

The interaction of the magnetic vector potential with a charged particle is $eA(x)$. In cgs units the classical field energy is

$$H_f^{\text{classical}} = \frac{1}{8\pi} \int_{\mathbb{R}^3} \{B(x)^2 + E(x)^2\} dx. \quad (63)$$

With $B(x) = \text{curl}A(x)$, we use the Coulomb (or radiation gauge) so that $\text{div}A(x) = 0$ and $\text{div}E(x) = 0$.

We define $a_\lambda(k)$ and its complex-conjugate (classically) or adjoint (quantum-mechanically), $a_\lambda^*(k)$, in terms of the Fourier transform of (the real fields) $A(x)$ and $E(x)$ as follows.

$$A(x) = \frac{\sqrt{\hbar c}}{2\pi} \sum_{\lambda=1}^2 \int_{\mathbb{R}^3} \frac{\varepsilon_\lambda(k)}{\sqrt{|k|}} \left(a_\lambda(k) e^{ik \cdot x} + a_\lambda^*(k) e^{-ik \cdot x} \right) dk, \quad (64)$$

$$E(x) = \frac{i\sqrt{\hbar c}}{2\pi} \sum_{\lambda=1}^2 \int_{\mathbb{R}^3} \sqrt{|k|} \varepsilon_\lambda(k) \left(a_\lambda(k) e^{ik \cdot x} - a_\lambda^*(k) e^{-ik \cdot x} \right) dk, \quad (65)$$

in terms of which

$$B(x) = \frac{i\sqrt{\hbar c}}{2\pi} \sum_{\lambda=1}^2 \int_{\mathbb{R}^3} \frac{k \wedge \varepsilon_\lambda(k)}{\sqrt{|k|}} \left(a_\lambda(k) e^{ik \cdot x} - a_\lambda^*(k) e^{-ik \cdot x} \right) dk. \quad (66)$$

The parameter $\sqrt{\hbar c}/2\pi$ in (64–66) were chosen purely for convenience later on. The two unit vectors here, $\varepsilon_\lambda(k)$, $\lambda = 1, 2$, are perpendicular to each other and to k (which guarantees that $\text{div}A = 0$). They cannot be defined on the whole of $\mathbb{R}^3$ as smooth functions of k (although they can be so defined with the use of “charts”), but that will be of no concern to us.

Thus, when (64–66) are substituted in (63) we obtain (using Parseval's theorem and $\int \exp(ik \cdot x) dx = (2\pi)^3 \delta(k)$ and $|k|^2 \varepsilon_\lambda(k) \cdot \varepsilon_\lambda(-k) = -(k \wedge \varepsilon_\lambda(k)) \cdot (-k \wedge \varepsilon_\lambda(-k))$),

$$H_f^{\text{classical}} = \frac{1}{2} \hbar c \sum_{\lambda=1}^2 \int_{\mathbb{R}^3} |k| \{ a_\lambda^*(k) a_\lambda(k) + a_\lambda(k) a_\lambda^*(k) \} dk. \quad (67)$$

(Although $a_\lambda^*(k) a_\lambda(k) = a_\lambda(k) a_\lambda^*(k)$ for functions, this will not be so when $a_\lambda(k)$ is an operator. The form in (67) is that obtained after the substitution just mentioned.)

To complete the picture, we quantize the fields by making the $a_\lambda(k)$ into operators with the following commutation relations:

$$[a_\lambda(k), a_{\lambda'}^*(k')] = \delta_{\lambda, \lambda'} \delta(k - k') \quad \text{and} \quad [a_\lambda(k), a_{\lambda'}(k')] = 0. \quad (68)$$

The quantized field energy is obtained from (67, 68) and is given by the Hamiltonian operator

$$H_f = \hbar c \sum_{\lambda=1}^2 \int_{\mathbb{R}^3} |k| a_\lambda^*(k) a_\lambda(k) dk. \quad (69)$$

It agrees with (67) up to an additive “infinite constant”.

In the rest of this paper we omit $\hbar c$ since we use units in which $\hbar = c = 1$.

B. Appendix: Field Energy Bound

In this appendix we prove (79) which relates the localized classical field energy to the quantized field energy. A proof was given in [4]. The small generalization given here is a slightly modified version of that in [13, 14].

Consider a collection of operators (field modes), parametrized by $y \in \mathbb{R}^3$, and by j in some set of integers ($j \in \{1, 2, 3\}$ in our case of interest) given, formally, by

$$L_j(y) = \sum_{\lambda=1}^2 \int_{|k| < \Lambda} \sqrt{|k|} \widehat{v}_{\lambda,j}(k) e^{ik \cdot y} a_\lambda(k) dk, \quad (70)$$

where $\widehat{v}_{\lambda,j}$ is the Fourier transform of some arbitrary complex function $v_{\lambda,j}(x)$. Our convention for the Fourier transform of a general function $g(x)$ is

$$\widehat{g}(k) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} g(x) e^{+ik \cdot x} dx \quad \text{and} \quad g(x) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \widehat{g}(k) e^{-ik \cdot x} dk. \quad (71)$$

The following lemma is elementary. It involves $v_{\lambda,j}(x)$ and a summable function $w(x)$, with a norm defined by

$$\|w\|_v := \sup_{f_\lambda(x)} \frac{\sum_j \int_{\mathbb{R}^3} |\sum_\lambda f_\lambda * v_{\lambda,j}(x)|^2 |w(x)| dx}{\sum_\lambda \int_{\mathbb{R}^3} |f_\lambda(x)|^2 dx}, \quad (72)$$

where $*$ is convolution.

Lemma B.1 (Lower bound on field energy). *Assume that $\|w\|_v \leq 1$. Then*

$$H_f \geq \sum_j \int_{\mathbb{R}^3} w(y) L_j^*(y) L_j(y) dy. \quad (73)$$

Moreover, if $w(y) \geq 0$, for all y then

$$\begin{aligned} H_f &\geq \frac{1}{4} \sum_j \pm \int_{\mathbb{R}^3} w(y) (L_j(y) \pm L_j^*(y))^2 dy \\ &\quad - \frac{1}{2} \sum_j \sum_{\lambda=1}^2 \int_{|k| < \Lambda} |k| |\widehat{v}_{\lambda,j}(k)|^2 dk \int_{\mathbb{R}^3} w(y) dy, \end{aligned} \quad (74)$$

for any choice of $+$ or $-$ for each j . (Note that $-(L - L^*)^2 \geq 0$.)

Proof. The difference of the two sides in (73) is a quadratic form of the type $\sum_{\lambda,\lambda'} \int \int a_{\lambda}^*(k) Q(k, \lambda : k', \lambda') a_{\lambda'}(k') dk dk'$. In order to establish (73) it is necessary and sufficient to prove that the matrix $Q(k, \lambda : k', \lambda')$ is positive semidefinite. This is the condition that

$$\sum_{\lambda=1}^2 \int_{|k|<\Lambda} |\widehat{f}_{\lambda}(k)|^2 dk \geq (2\pi)^{3/2} \sum_j \sum_{\lambda,\lambda'=1}^2 \int \int_{|k|,|k'|<\Lambda} \overline{\widehat{f}_{\lambda}(k) \widehat{v}_{\lambda,j}(k)} \widehat{f}_{\lambda'}(k') \widehat{v}_{\lambda',j}(k') \widehat{w}(k' - k) dk dk' \quad (75)$$

for all L^2 functions $\widehat{f}_{\lambda}(k)$. Condition (75) is just the condition that $\|w\|_v \leq 1$, since $\widehat{f * v}(k) = (2\pi)^{3/2} \widehat{f}(k) \widehat{v}(k)$.

To obtain (74) from (73) we use the three facts that $w(x) \geq 0$, that

$$L_j^*(y) L_j(y) = L_j(y) L_j^*(y) - \sum_{\lambda=1}^2 \int_{|k|<\Lambda} |k| |\widehat{v}_{\lambda,j}(k)|^2 dk, \quad (76)$$

and that, quite generally for operators,

$$\pm (LL + L^*L^*) \leq L^*L + LL^*. \quad (77)$$

□

The following examples are important. First, we define the ultraviolet cutoff fields A^{Λ} , B^{Λ} , E^{Λ} as in (64,66,65) except that the k integration is over $|k| \leq \Lambda$ instead of $\mathbb{R}^3$. E.g.,

$$B^{\Lambda}(x) = \frac{i}{2\pi} \sum_{\lambda=1}^2 \int_{|k|\leq\Lambda} \frac{k \wedge \varepsilon_{\lambda}(k)}{\sqrt{|k|}} \left(a_{\lambda}(k) e^{ik \cdot x} - a_{\lambda}^*(k) e^{-ik \cdot x} \right) dk, \quad (78)$$

recalling that $\hbar = c = 1$.

This notation, A^{Λ} , B^{Λ} , E^{Λ} , with the superscript Λ , will be used in this appendix only.

For the first two examples we define $\widehat{v}_{\lambda,j}(k) = ((k \wedge \varepsilon_{\lambda}(k))_j / (2\pi)^{3/2} |k|)$, so that $i(L_j(x) - L_j^*(x)) = B_j^{\Lambda}(x) / \sqrt{2\pi}$ for $j = 1, 2, 3$.

Example 1. Assume that $0 \leq w(y) \leq 1$ for all y . Then Lemma B.1 implies

$$H_f \geq \frac{1}{8\pi} \int_{\mathbb{R}^3} B^{\Lambda}(x)^2 w(x) dx - \frac{\Lambda^4}{8\pi^2} \int_{\mathbb{R}^3} w(y) dy$$

 . \quad (79)

To verify the norm condition (75) we note first that in view of (72) it suffices to assume that $w(y) \equiv 1$. Then, $\widehat{w}(k) = (2\pi)^{3/2} \delta(k)$. On the right side of (75) we may use the equality $\sum_j \widehat{v}_{\lambda,j}(k) \widehat{v}_{\lambda',j}(k) = (2\pi)^{-3} \delta_{\lambda,\lambda'}$ (because $((k \wedge \varepsilon_{\lambda}(k))/|k|)_j$ are the three components of two orthonormal vectors). Thus, (75) is not only satisfied, it is also an identity with this choice of $\widehat{w}$. Finally, (74) is exactly (79) since $\int_{|k|\leq\Lambda} |k| dk = \pi \Lambda^4$.

Example 2. Take $w(y) = C\delta(x - y)$, with x some fixed point in $\mathbb{R}^3$ and where $C > 0$ is some constant. Then $\widehat{w}(k) = C(2\pi)^{-3/2} \exp(ik \cdot x)$ and $|\widehat{w}(k)| = C(2\pi)^{-3/2}$. We take $\widehat{v}_{\lambda, j}(k)$ as in Example 1. The right side of condition (75) equals $C \sum_j \int_{|k| \leq \Lambda} \sum_{\lambda} \widehat{f}_{\lambda}(k) \widehat{v}_{\lambda, j}(k) \exp(ik \cdot x) dk|^2$, and a simple variational argument then says that we should choose $\widehat{f}_{\lambda}(k) = \exp(-ik \cdot x) \overline{\widehat{v}_{\lambda}(k)} \cdot V$, where V is some fixed vector. By spherical symmetry we can take $\widehat{f}_{\lambda}(k) = \widehat{v}_{\lambda, 1}(k)$, and (75) becomes

$$1 \geq C \int_{|k| \leq \Lambda} \sum_{\lambda} |\widehat{v}_{\lambda, 1}(k)|^2 dk = C \frac{2}{3} \cdot \frac{1}{(2\pi)^3} \cdot \frac{4\pi}{3} \Lambda^3. \quad (80)$$

With $C = 9\pi^2 \Lambda^{-3}$, and with

$$\sum_{\lambda} \sum_j \int_{|k| < \Lambda} |k| |\widehat{v}_{\lambda, j}(k)|^2 dk \int_{\mathbb{R}^3} w(y) dy = 2 \cdot (2\pi)^{-3} \cdot (4\pi/4) \Lambda^4 C,$$

(74) becomes $H_f \geq \frac{1}{4} \cdot C \cdot (2\pi)^{-1} B^{\Lambda}(x)^2 - (1/8)\pi^{-2} \cdot \Lambda^4 \cdot C$, or

$$\boxed{H_f \geq \frac{9\pi}{8} \Lambda^{-3} B^{\Lambda}(x)^2 - \frac{9}{8} \Lambda}. \quad (81)$$

This can also be used [13, 14] with x being the electron coordinate (which is an operator, to be sure, but is one that commutes with the field operators).

Example 3. If we replace $(k \wedge \varepsilon_{\lambda}(k))_j / |k|$ in $v_{\lambda, j}(k)$ by $(\varepsilon_{\lambda}(k))_j$ then everything goes through as before and we obtain (79) and (81) with $E^{\Lambda}(x)^2$ in place of $B^{\Lambda}(x)^2$.

Example 4. We now take $\widehat{v}_{\lambda, j}(k) = (\varepsilon_{\lambda}(k))_j (2\pi)^{-3/2} / |k|$, so that $(L_j(x) + L_j^*(x)) = A_j^{\Lambda}(x) / \sqrt{2\pi}$ for $j = 1, 2, 3$. The analysis proceeds as in Example 2, except that the normalization condition (80) becomes

$$1 \geq C \int_{|k| \leq \Lambda} \sum_{\lambda} |v_{\lambda, 1}(k)|^2 dk = C \frac{2}{3} \cdot \frac{1}{(2\pi)^3} \cdot 4\pi \Lambda$$

which leads to $C = 3\pi^2 \Lambda^{-1}$. We also have

$$\sum_{\lambda} \sum_j \int_{|k| < \Lambda} |k| |\widehat{v}_{\lambda, j}(k)|^2 dk \int_{\mathbb{R}^3} w(y) dy = 2 \cdot (2\pi)^{-3} \cdot (4\pi/2) \Lambda^2 C,$$

so that (74) becomes

$$\boxed{H_f \geq \frac{3\pi}{8} \Lambda^{-1} A^{\Lambda}(x)^2 - \frac{3}{4} \Lambda}. \quad (82)$$

C. Appendix: Spectral Properties of the Dirac Operators

In this appendix we sketch a proof of the fact that the operators $D_i(A)$ commute in the sense that all their spectral projections commute. First we start with some remarks concerning the self-adjointness of

$$D_i(A) = \alpha_i \cdot (-i\nabla_i + \sqrt{\alpha}A(x_i)) + m\beta. \quad (83)$$

The subscript after α is a reminder that the matrix acts on the spinor associated with the i^{th} particle. It is not easy to characterize the domain for this operator, but it is certainly defined and symmetric on $\mathcal{H}_N^0 := C_c^\infty(\mathbb{R}^3) \otimes \mathcal{F}_{\text{finite}}$, where $\mathcal{F}_{\text{finite}}$ denotes the vectors in Fock space with finitely many photons. This is a dense subset of $\mathcal{H}_N$. We shall show that $D_i(A)$ is essentially self-adjoint on $\mathcal{H}_N^0$. To prove this we resort to a version of Nelson's commutator theorem given in [22], Theorem X.37.

Define the operator

$$\nu = 1 + \sum_{i=1}^N (-\Delta_i + m^2) + \Lambda H_f + \Lambda^2. \quad (84)$$

Observe that $\sum_{i=1}^N (-\Delta_i + m^2)$ acts as a multiplication operator on Fourier space and H_f acts on the n photon component $\Phi_n(x_1, \dots, x_N; \tau_1, \dots, \tau_N; k_1, \dots, k_n, \lambda_1, \dots, \lambda_n)$ by multiplication with $\sum_{i=1}^n \omega(k_i)$. The domain $\mathcal{H}_N^0$ is a domain of essential self-adjointness for ν . Certainly, $\nu \geq 1$ as an operator.

We shall show that there exists a constant c such that for all $\Psi \in \mathcal{H}_N^0$,

$$\|D_i(A)\Psi\| \leq c\|\nu\Psi\|. \quad (85)$$

Certainly,

$$\|D_i(A)\Psi\| \leq \left(\Psi, (-\Delta_i + m^2)\Psi \right)^{1/2} + \left(\Psi, A(x_i)^2\Psi \right)^{1/2}, \quad (86)$$

and by Example 4 in Appendix B,

$$A(x)^2 \leq \frac{8}{3\pi} \Lambda H_f + \frac{2}{\pi} \Lambda^2. \quad (87)$$

The estimate

$$\|D_i(A)\Psi\| \leq c\|\nu\Psi\| \quad (88)$$

follows easily from this.

Next, we show that there exists a constant d such that for all $\Psi \in \mathcal{H}_N^0$,

$$|(D_i(A)\Psi, \nu\Psi) - (\nu\Psi, D_i(A)\Psi)| \leq d\|\nu^{1/2}\Psi\|^2. \quad (89)$$

Since $-i\alpha_i \cdot \nabla_i + m\beta$ commutes with ν when applied to vectors in $\mathcal{H}_N^0$, the above estimate reduces to

$$|(\alpha_i \cdot A(x_i)\Psi, \nu\Psi) - (\nu\Psi, \alpha_i \cdot A(x_i)\Psi)| \leq d\|\nu^{1/2}\Psi\|^2, \quad (90)$$

where we have dropped the fine structure constant. Since ν as well as $A(x)$ preserve $\mathcal{H}_N^0$ and are symmetric, we can rewrite the above inequality as

$$|(\Psi, [\alpha_i \cdot A(x_i), \nu] \Psi)| \leq d \|\nu^{1/2} \Psi\|^2. \quad (91)$$

The commutator is the sum of

$$[\alpha_i \cdot A(x), -\Delta] = i \sum_{j=1}^3 \alpha_i \cdot (\partial_j A(x) \partial_j + \partial_j \partial_j A(x)) =: X, \quad (92)$$

and the operator

$$-i\Lambda \alpha_i \cdot E(x_i) = [\alpha_i \cdot A(x_i), \Lambda H_f], \quad (93)$$

where $E(x)$ is the electric field (65). By Schwarz's inequality,

$$|(\Psi, X\Psi)| \leq \left((\Psi, (\nabla A)^2 \Psi) + (\Psi, -\Delta \Psi) \right), \quad (94)$$

and

$$\Lambda |(\Psi, E(x)\Psi)| \leq \Lambda \left(\|\Psi\|^2 + (\Psi, E(x)^2 \Psi) \right). \quad (95)$$

By Example 3 in Appendix B, it follows that as quadratic forms

$$E(x)^2 \leq \frac{8}{9\pi} \Lambda^3 H_f + \frac{1}{\pi} \Lambda^4, \quad (96)$$

$$(\nabla A)(x)^2 \leq \frac{8}{9\pi} \Lambda^3 H_f + \frac{1}{\pi} \Lambda^4. \quad (97)$$

The last inequality does not appear in Appendix B exactly as stated, but it can be derived in precisely the same fashion as the one for the magnetic field displayed there. The estimates (94)-(97) yield (provided $\Lambda \geq 1$)

$$|(D_i(A)\Psi, \nu\Psi) - (\nu\Psi, D_i(A)\Psi)| \quad (98)$$

$$\leq C \left((\Psi, (\Lambda^3 H_f + \Lambda^4) \Psi) + (\Psi, (-\Delta + \Lambda) \Psi) \right) \quad (99)$$

$$\leq C \Lambda^2 (\Psi, \nu\Psi), \quad (100)$$

for some constant C , which is the desired estimate.

Thus, the operator $D_i(A)$ is essentially self-adjoint on $\mathcal{H}_N^0$. This operator, being a sum of two self-adjoint operators, $U_i := -i\alpha_i \cdot \nabla_i + m\beta$ and $V_i := \alpha_i \cdot \sqrt{\alpha} A(x_i)$, is naturally defined on $\mathcal{D}(U_i) \cap \mathcal{D}(V_i)$ and is symmetric there. Since, $\mathcal{H}_N^0 \subset \mathcal{D}(U_i) \cap \mathcal{D}(V_i)$ we also know that $D_i(A)$ is essentially self-adjoint on $\mathcal{D}(U_i) \cap \mathcal{D}(V_i)$. Thus, by Theorem VIII.31 in [21] the Trotter product formula is valid, i.e.

$$e^{itD_i(A)} = s - \lim_{m \rightarrow \infty} T_i(t/m)^m, \quad (101)$$

where

$$T_i(t/m) := e^{i(t/m)U_i} e^{i(t/m)V_i}. \quad (102)$$

Certainly, the operator e^{itU_i} commutes with e^{isU_j} and e^{isV_j} , and likewise e^{itV_i} commutes with e^{isU_j} and e^{isV_j} for all $j \neq i$, and hence $T_i(t/m)^m$ commutes with $T_j(s/n)^n$ for all s, t, m and n .

We shall use this to show the following

Lemma C.1. *For any two real numbers s and t the unitary groups $e^{itD_i(A)}$ and $e^{isD_j(A)}$ commute. Moreover, this implies that the spectral projections associated with $D_i(A)$ and $D_j(A)$ commute.*

Proof. For $\Psi \in \mathcal{H}_N$,

$$\|e^{itD_i(A)}e^{isD_j(A)}\Psi - e^{isD_j(A)}e^{itD_i(A)}\Psi\| \quad (103)$$

$$= \lim_{m \rightarrow \infty} \|T_i(t/m)^m e^{isD_j(A)}\Psi - e^{isD_j(A)}T_i(t/m)^m \Psi\| \quad (104)$$

$$= \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \|T_i(t/m)^m T_j(s/n)^n \Psi - T_j(s/n)^n T_i(t/m)^m \Psi\| = 0. \quad (105)$$

The statement about the spectral projections follows from Theorem VIII.13 in [21]. $\square$

D. Appendix: Projections and Symmetries

The difficulty in defining the physical space $\mathcal{H}_N^{\text{phys}}$ comes from the fact that the projection onto the positive energy subspace acts also on the Fock space. This is in contrast to [17] where no such problem arises. There the action of the permutation group obviously commuted with the projection onto the positive energy subspace. In our more general setting the commutation is still true but an explanation is needed, which we try to give with a minimal amount of formality.

First, consider the one particle space $\mathcal{H}_1 = L^2(\mathbb{R}^3; \mathbb{C}^4) \otimes \mathcal{F}$. The Dirac operator, as shown in Appendix C, is a self-adjoint operator on $\mathcal{H}_1$ and we denote its projections onto the positive and negative energy subspace by P^+ and P^- . Note that $P^+ + P^-$ is the identity. As explained in Sect. 4, the two projections are unitarily equivalent via

$$P^- = U^* P^+ U = -U P^+ U, \quad (106)$$

where, as in (39),

$$U = \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix}. \quad (107)$$

The projection onto the positive energy subspace associated with the Dirac operator $D_i(A)$ is defined in the following fashion. Consider the vector Ψ as in (12) and fix $N-1$ x 's and τ 's, namely all those except x_i and τ_i . For almost every such choice (with respect to Lebesgue measure) the vector Ψ defines a vector in $\mathcal{H}_1$. We know how the $P^\pm$ act on such a vector and the extension to $\mathcal{H}_N$ we denote by $P_i^\pm \Psi$. It was shown in Appendix C that these spectral projectors commute with each other.

Other interesting operators on $\mathcal{H}_N$ are the permutations. A permutation $\text{Per}_{1,2}$, for example, just exchanges the electron labels 1 and 2. From what has been explained above we have the formula

$$\text{Per}_{1,2} P_1^\pm = P_2^\pm \text{Per}_{1,2}. \quad (108)$$

An immediate consequence is that $P^\pm := \prod_{i=1}^N P_i^\pm$ commutes with permutations.

From this it follows that (14) can be rewritten as

$$\mathcal{H}_N^{\text{phys}} := \mathcal{A} P^+ \mathcal{H}_N = P^+ \mathcal{A} \mathcal{H}_N. \quad (109)$$

We now address the question whether $\mathcal{H}_N^{\text{phys}}$ is trivial or not. Denote by $\mathcal{K}$ the subspace of $\mathcal{H}_N$ which consists of antisymmetric vectors Ψ with the property that $U_j \Psi = i \Psi$ for $j = 1, \dots, N$. The operators U_j are defined in (39). Certainly each U_j has eigenvalues i and $-i$. The space $\mathcal{K}$ is certainly infinite dimensional. It contains, e.g., determinantal vectors in $L^2(\mathbb{R}^3; \mathbb{C}^4)$ tensor the photon vacuum.

Lemma D.1 ($\mathcal{H}_N^{\text{phys}}$ is large). *The space $\mathcal{H}_N^{\text{phys}}$ is infinite dimensional.*

Proof. We shall show that $2^{N/2} P^+$ is an isometry from $\mathcal{K}$ into $\mathcal{H}_N^{\text{phys}}$. Let I be a subset of the integers $\{1, \dots, N\}$ and let J be its complement. Let

$$P_I = \Pi_{i \in I} P_i^- \Pi_{j \in J} P_j^+. \quad (110)$$

Note that $\sum_I P_I = \text{identity}$. Note also that

$$P_I = (-1)^{|I|} U_I \Pi_{i \in I} P_i^+ U_I \Pi_{j \in J} P_j^+, \quad (111)$$

which implies that $\|P_I \Psi\| = \|P^+ \Psi\|$. This shows in particular that

$$\|P^+ \Psi\|^2 = 2^{-N} \|\Psi\|^2, \quad (112)$$

which proves the isometry. $\square$

Since we always consider the symmetric operator H_N^{phys} in the sense of quadratic forms, it is necessary to construct a domain, $\mathcal{Q}_N$ that is dense in $\mathcal{H}_N^{\text{phys}}$ and on which every term in H_N^{phys} has a finite expectation value. Once it is shown that the quadratic form associated with H_N^{phys} is bounded below, it is closable and its closure defines a self-adjoint operator, the Friedrichs extension of H_N^{phys} .

We first start with a technical lemma that will allow us to approximate vectors in $\mathcal{H}_N^{\text{phys}}$.

Lemma D.2. *For any $\widehat{f}$ with*

$$C_f := \int_{-\infty}^{\infty} |\widehat{f}(t)|(1 + |t|) dt < \infty \quad (113)$$

and

$$f(D(A)) = \int e^{-itD(A)} \widehat{f}(t) dt, \quad (114)$$

we have that

$$\|\sqrt{1 + H_f} f(D(A)) \Psi\| \leq \max\{\sqrt{1 + 9\Lambda/8}, \sqrt{8/9\pi} \Lambda^{3/2}\} C_f \|\sqrt{1 + H_f} \Psi\|, \quad (115)$$

for all $\Psi \in \mathcal{H}_1$.

Proof. We shall assume that Ψ is normalized. Since

$$\|\sqrt{1 + H_f} f(D(A)) \Psi\| \leq \int |\widehat{f}(t)| \|\sqrt{1 + H_f} e^{-itD(A)} \Psi\| dt, \quad (116)$$

it suffices to prove the estimate

$$K(t) := \|\sqrt{1 + H_f} e^{-itD(A)} \Psi\| \leq C(1 + |t|) \|\sqrt{1 + H_f} \Psi\|. \quad (117)$$

A simple calculation yields

$$\begin{aligned} \frac{d}{dt} K^2(t) &= \left(e^{-itD(A)} \Psi, i[H_f, D(A)] e^{-itD(A)} \Psi \right) \\ &= - \left(e^{-itD(A)} \Psi, \alpha \cdot E(x) e^{-itD(A)} \Psi \right). \end{aligned} \quad (118)$$

Here $E(x)$ is the electric field

$$E(x) = \frac{i}{2\pi} \sum_{i=1}^2 \int_{|k| \leq \Lambda} dk \varepsilon_\lambda(k) \sqrt{\omega(k)} \left(e^{ik \cdot x} a_\lambda(k) - e^{-ik \cdot x} a_\lambda^*(k) \right). \quad (119)$$

By Schwarz's inequality

$$\frac{d}{dt} K^2(t) \leq \left(e^{-itD(A)} \Psi, E(x)^2 e^{-itD(A)} \Psi \right)^{1/2}. \quad (120)$$

By Example 3 in Sect. B

$$E(x)^2 \leq \frac{8}{9\pi} \Lambda^3 H_f + \frac{1}{\pi} \Lambda^4, \quad (121)$$

and hence $K^2(t)$ satisfies the differential inequality

$$\frac{d}{dt} K^2(t) \leq \left(A K^2(t) + B \right)^{1/2}, \quad (122)$$

where $A = \frac{8}{9\pi} \Lambda^3$ and $B = \frac{1}{\pi} \Lambda^4$. This can be readily solved (using $K(t) \geq 1$) to yield the estimate

$$K(t) \leq (1 + B/A)^{1/2} K(0) + \sqrt{A} t. \quad (123)$$

Thus

$$\|\sqrt{1 + H_f} e^{-itD(A)} \Psi\| \leq C(1 + |t|) \|\sqrt{1 + H_f} \Psi\|, \quad (124)$$

where C is the maximum of $(1 + B/A)^{1/2}$ and $\sqrt{A}$. $\square$

Next we consider a sequence of functions $f_n \in C_c^\infty((0, \infty))$ everywhere less than or equal to 1, such that f_n is identically equal to 1 on the interval $[1/n, n]$. Clearly, as $n \rightarrow \infty$, $\Pi_{i=1}^N f_n(D_i(A)) \rightarrow P^+$ strongly in $\mathcal{H}_N$ and hence $\mathcal{A} \Pi_{i=1}^N f_n(D_i(A)) \rightarrow \mathcal{A} P^+$ strongly in $\mathcal{H}_N$. We denote the range of $\mathcal{A} \Pi_{i=1}^N f_n(D_i(A))$ restricted to the subspace of $\mathcal{H}_N$ consisting of states with *finite field energy expectation* by $\mathcal{Q}_N^n$. Finally we define the domain $\mathcal{Q}_N = \bigcup_{n=1}^\infty \mathcal{Q}_N^n$. Together with Lemma D.2 we have the following corollary.

Corollary D.3. *The domain $\mathcal{Q}_N$ is dense in $\mathcal{H}_N^{\text{phys}}$. Moreover for any vector $\Psi \in \mathcal{Q}_N$ the field energy H_f has finite expectation value.*

Proof. Simply note that the functions f_n have a rapidly decaying Fourier transform for each n . Therefore, by Lemma D.2 the field energy has a finite expectation value for any vector $\Psi \in \mathcal{Q}_N^n$. Note, as before, the antisymmetrization operator $\mathcal{A}$ commutes with $\Pi_{i=1}^N f_n(D_i(A))$. Thus, the field energy has finite expectation value for any $\Psi \in \mathcal{Q}_N$. The density of $\mathcal{Q}_N$ in $\mathcal{H}_N^{\text{phys}}$ was shown before. $\square$

Now we are ready to state the main lemma of this section.

Lemma D.4. *For every $\Psi \in \mathcal{Q}_N$, the Dirac operators $D_i(A)$, the Coulomb potential V_c and the field energy H_f have finite expectation values. Thus, H_N^{phys} is defined as a quadratic form on $\mathcal{Q}_N$ which is dense in $\mathcal{H}_N^{\text{phys}}$.*

Proof. The operators $D_i(A)^2$ have finite expectation values on $\mathcal{Q}_N$. They are of the form $\widehat{T}^P(A) = [(p + \sqrt{\alpha}A(x))^2 + \sqrt{\alpha}\sigma \cdot B(x)] \otimes I_2$, where I_2 is the 2×2 identity. By (81) the magnetic field is bounded by the field energy and hence has finite expectation value on the domain $\mathcal{Q}_N$. Thus, $(p_i + \sqrt{\alpha}A(x_i))^2 \otimes I_2$ also has finite expectation value on $\mathcal{Q}_N$ for $i = 1, \dots, N$, and hence the Coulomb potential V_c , which is relatively bounded with respect to $\sum_{i=1}^N (p_i + \sqrt{\alpha}A(x_i))^2$ has finite expectation values on $\mathcal{Q}_N$. $\square$

E. Appendix: Various Forms of Instability

In the introduction we talked about the need of using the positive spectral subspace of the Dirac operator $D(A)$, which includes the magnetic vector potential; this led to all sorts of complications in the analysis leading to our main stability Theorem 2.1. In this section we show that various models in which an electron is defined, instead, by the positive spectral subspace of the *free* Dirac operator $D(0)$ are unstable. In the case of a classical magnetic field such an analysis was carried out in [17] and greatly simplified in [9]. Also, in [9] the stability analysis was carried out for a quantized radiation field without a cutoff. In what follows, we rely mostly on the work in [9]. We also show that the $D(A)$ choice is unstable if $Z\alpha$ or α is too large – as expected.

All the results about stability and instability are summarized in the two tables in Sect. 1. We remind the reader that instability of the first kind means that the Hamiltonian is unbounded below, while instability of the second kind means that it is bounded below but not by a constant times $N + K$.

E.1. Instability without Coulomb potential (free Dirac operator). Already the free problem, i.e., without Coulomb interactions, shows signs of instability. The Hamiltonian is given by

$$H_N^q = \sum_{j=1}^N D_j(A) + H_f. \quad (125)$$

If the field is classical, H_f has to be replaced by $(1/8\pi) \int_{\mathbb{R}^3} |B(x)|^2 dx$ as in (1).

We consider first the case where the magnetic vector potential is classical. In particular the Hilbert space $\mathcal{H}_N^{\text{free}}$ is the antisymmetric tensor product of N copies of $P^+L^2(\mathbb{R}^3; \mathbb{C}^4)$, i.e., the part of $L^2(\mathbb{R}^3; \mathbb{C}^4)$ that is in the positive spectral subspace of the *free* Dirac operator. Note that there is no Fock space in this case. In [9] Theorems 1 and 3 the authors construct, for any N , a trial Slater determinant ψ in $\mathcal{H}_N^{\text{free}}$, and a classical field A so that the energy is bounded above by

$$\left(\psi, H_N^{\text{classical}} \psi \right) =: \mathcal{E}(\psi, A) \leq aN^{4/3} - \alpha bN^2, \quad (126)$$

where a and b are constants independent of N . The scaling

$$\psi \rightarrow \psi_\mu, \text{ and } A \rightarrow A^\mu, \tag{127}$$

where

$$\psi_\mu(x_1, \dots, x_N; \tau_1, \dots, \tau_N) = \mu^{3N/2} \psi(\mu x_1, \dots, \mu x_N; \tau_1, \dots, \tau_N) \tag{128}$$

and

$$A^\mu(x) = \mu A(\mu x) \tag{129}$$

can be used to get the upper bound

$$\mathcal{E}(\psi_\mu, A^\mu) \leq \mu \left(aN^{1/3} - \alpha bN^2 \right). \tag{130}$$

Thus, by choosing $N > a\alpha^{-3/2}/b$, the ground state energy is negative and can be driven to $-\infty$ by letting $\mu \rightarrow \infty$, i.e., *stability of the first kind is violated*.

Using coherent states, it was shown in [9] that this same result extends to the problem with a quantized magnetic vector potential without ultraviolet cutoff.

If $A(x)$ carries an ultraviolet cutoff, the μ scaling argument cannot be applied. The energy, however, is not bounded below by $\text{const.} \times N$, as we see from (126), and hence *stability of the second kind* is also violated – and this no matter how small α may be and whether the magnetic vector potential is quantized or not. The reader might wonder how to construct an $A(x)$ that satisfies the conditions in [9] and, at the same time has an ultraviolet cutoff Λ . Remark 1 on p. 1782 of [9] explains that almost any cutoff $A(x)$ will suffice for the purpose.

Notice that when we use the positive spectral subspace of $D(A)$ instead, the stability of the problem without Coulomb potential is completely trivial, since the Hamiltonian is positive, by definition.

E.2. Instability with Coulomb potential (free Dirac operator). Adding the Coulomb potential complicates the analysis owing to the repulsion between the electrons which is present even if there are no nuclei. To some extent this positive energy is balanced by the electron-nuclei attraction if sufficiently many nuclei with sufficiently strong charges are present. It is shown in [17] that if

$$\sum_{j=1}^K Z_j \geq (\text{const.})\alpha^{-3/2} \text{ and } \sum_{j=1}^K Z_j^2 \geq 2, \tag{131}$$

then the positions of the nuclei can be chosen such that the total Coulomb energy is negative. Thus, if in addition, $N\alpha^{3/2}$ is sufficiently large, *stability of the first kind* does not hold for classical magnetic vector potentials as well as for a quantized magnetic vector potential (without ultraviolet cutoff) – no matter how small α may be.

The situation is more complicated when the field carries an ultraviolet cutoff. The main reason is that the field variable is no longer an active participant for driving the energy towards minus infinity, but it is an active participant in destroying stability of the second kind. We have

Lemma E.1. *Let $\alpha > 0$ and assume that (131) holds. Then the system using the projection onto the positive subspace of the free Dirac operator $D(0)$ is unstable of the second kind, even with an ultraviolet cutoff.*

Proof. The lemma follows immediately from (130) and (131) together with the observation in [9] on how to use coherent states to carry these results over to the quantized field case. $\square$

Lemma E.1 is the main reason that the restriction to the positive spectral subspace of $D(0)$ is inadequate for a model of matter interacting with radiation.

The main result of this paper is the stability of the second kind for the system where the positive subspace of the Dirac operator $D(A)$ is used. This result holds provided that $\max_j Z_j \alpha$ and α is sufficiently small. Our final Lemma E.2 shows that that the two conditions $\max_j Z_j \alpha$ small and α small are in fact necessary. It suffices to show this for the case of one electron interacting with K nuclei, each having charge Z . We have to assume that the electron mass is strictly positive, but we suspect that this assumption is technical and not needed.

Lemma E.2. *Assume that $Z\alpha > 4/\pi$ and $m > 0$. Then the one-electron Hamiltonian*

$$H_1^{\text{phys}} = P^+ \left(D(A) - Z\alpha \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} + H_f \right) P^+ \quad (132)$$

is unbounded below. Moreover, there is a number α_c such that for any fixed $\alpha > \alpha_c$ and any fixed $Z > 0$, there exists K sufficiently large so that this Hamiltonian is unbounded below.

For the classical instead of the quantized A field see [6, 10, 1, 24, 25].

Proof. The idea is to reduce this problem to the relativistic one without spin and without radiation field. Set $D_+ = P^+ D(A)$ and $D_- = -P^- D(A)$, so that $D(A) = D_+ - D_-$ and $|D(A)| = D_+ + D_-$. As a trial function we pick $\psi = g \otimes |0\rangle$, where g is a spinor that satisfies $Ug = ig$, recalling the definition of U from Appendix D (106, 107), and where $|0\rangle$ is the photon vacuum. A straightforward calculation (which repeatedly uses Schwarz's inequality and the facts that $P^- = U^* P^+ U = -U P^+ U$, and hence $P^- \psi = -i U P^+ \psi$) shows that

$$\begin{aligned} & (\psi, H_1^{\text{phys}} \psi) \\ &= \left(P^+ \psi, \left[D_+ - Z\alpha \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} + H_f \right] P^+ \psi \right) \\ &= \frac{1}{2} \left(P^+ \psi, \left[D_+ - Z\alpha \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} \right] P^+ \psi \right) \\ &\quad + \frac{1}{2} \left(P^- \psi, \left[D_- - Z\alpha \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} \right] P^- \psi \right) \\ &\quad + (\psi, P^+ H_f P^+ \psi) \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{2} \left(P^+ \psi, \left[D_+ - \frac{Z\alpha}{2} \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} \right] P^+ \psi \right) \\ &\quad + \frac{1}{2} \left(P^- \psi, \left[D_- - \frac{Z\alpha}{2} \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} \right] P^- \psi \right) \\ &\quad - \operatorname{Re} \left(P^+ \psi, \frac{Z\alpha}{2} \sum_{j=1}^K \frac{1}{|x - R_j|} P^- \psi \right) + (\psi, P^+ H_f P^+ \psi) \\ &\leq \frac{1}{2} \left(\psi, \left[|D(A)| - \frac{Z\alpha}{2} \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} \right] \psi \right) \\ &\quad + (\psi, P^+ H_f P^+ \psi). \end{aligned}$$

The lemma will be proved by showing that the first term in the last expression can be made as negative as we like while the field energy term is uniformly bounded. Note, as in (9), that $|D(A)| = \sqrt{\widehat{T}^P(A) + m^2}$. The operator inequality

$$(p + \sqrt{\alpha} A(x))^2 \leq (1 + \varepsilon) p^2 + \left(1 + \frac{1}{\varepsilon}\right) \alpha A(x)^2 \tag{133}$$

follows easily from Schwarz’s inequality, for any $\varepsilon > 0$. From (82) we have that

$$A(x)^2 \leq \frac{8}{3\pi} \Lambda H_f + \frac{2}{\pi} \Lambda^2, \tag{134}$$

and from (81) we have that

$$B(x)^2 \leq \frac{8}{9\pi} \Lambda^3 H_f + \frac{1}{\pi} \Lambda^4. \tag{135}$$

Using the operator monotonicity of the square root it follows that

$$\sqrt{T^P(A) + m^2} \leq \sqrt{(1 + \varepsilon) p^2 + X_\varepsilon \Lambda H_f + Y_\varepsilon \Lambda^2 + m^2}, \tag{136}$$

where X_ε and Y_ε are constants that tend to infinity as ε tends to zero.

Thus, recalling that $\psi = g \otimes |0\rangle$,

$$(\psi, |D(A)| \psi) \leq \left(g, \sqrt{(1 + \varepsilon) p^2 + Y_\varepsilon \Lambda^2} g \right). \tag{137}$$

The remaining task is to analyze the quadratic form

$$\left(g, \left[\sqrt{(1 + \varepsilon) p^2 + Y_\varepsilon \Lambda^2 + m^2} - \frac{Z\alpha}{2} \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} \right] g \right). \tag{138}$$

For any fixed ε , Λ and m the terms $Y_\varepsilon \Lambda^2 + m^2$ can be scaled away, and this leads to the quadratic form

$$\left(g, \left[\sqrt{(1+\varepsilon)p^2} - \frac{Z\alpha}{2} \sum_{j=1}^K \frac{1}{|x - R_j|} + Z^2 \alpha \sum_{i < j} \frac{1}{|R_i - R_j|} \right] g \right), \quad (139)$$

which has been analyzed in detail. Kato [11] showed that instability of the first kind occurs if $Z\alpha/2\sqrt{1+\varepsilon} > 2/\pi$ which yields our first stated condition for instability. Later on, it was shown in [5] that there exists α_c so that for $\alpha > \alpha_c$ and for any $Z > 0$ there exists K so that an instability of the first kind occurs. See also [19], Theorem 3.

Next we address the field energy term and it is here where the assumption about the positive mass comes in. The projection P^+ can be written as $P^+ = (1/2)(I + D(A)/|D(A)|)$ and, since ψ is proportional to the vacuum,

$$\begin{aligned} (\psi, P^+ H_f P^+ \psi) &= \frac{1}{4} \left(\psi, \frac{D(A)}{|D(A)|} H_f \frac{D(A)}{|D(A)|} \psi \right) \\ &= \sum_{\lambda} \int \omega(k) \left\| \left[a_{\lambda}(k), \frac{D(A)}{|D(A)|} \right] \psi \right\|^2 dk. \end{aligned} \quad (140)$$

With the help of the expression

$$\frac{D(A)}{|D(A)|} = \frac{1}{\pi} \int_0^\infty \frac{D(A)}{t + D(A)^2} \frac{1}{\sqrt{t}} dt, \quad (141)$$

and the fact that $(t + D(A)^2)^{-1} = \frac{1}{2}(D(A) - i\sqrt{t})^{-1} + \frac{1}{2}(D(A) + i\sqrt{t})^{-1}$, the commutator in the last expression of (140) can be written as

$$\begin{aligned} \left[a_{\lambda}(k), \frac{D(A)}{|D(A)|} \right] &= -\frac{1}{2\pi} \int_0^\infty \frac{1}{D(A) - i\sqrt{t}} [a_{\lambda}(k), D(A)] \frac{1}{D(A) - i\sqrt{t}} \frac{1}{\sqrt{t}} dt \\ &\quad - \frac{1}{2\pi} \int_0^\infty \frac{1}{D(A) + i\sqrt{t}} [a_{\lambda}(k), D(A)] \frac{1}{D(A) + i\sqrt{t}} \frac{1}{\sqrt{t}} dt. \end{aligned} \quad (142)$$

and

$$[a_{\lambda}(k), D(A)] = \frac{\varepsilon_{\lambda}(k)}{\sqrt{\omega(k)}} e^{-ik \cdot x} \chi_{\Lambda}(k),$$

where $\chi_{\Lambda}(k)$ denotes the characteristic function of the ball in k space that has radius Λ and is centered at the origin. Hence,

$$\begin{aligned} \left\| \left[a_{\lambda}(k), \frac{D(A)}{|D(A)|} \right] \psi \right\| &\leq \frac{1}{\sqrt{\omega(k)}} \frac{1}{\pi} \int_0^\infty \frac{1}{m^2 + t} \frac{1}{\sqrt{t}} dt \|\psi\| \chi_{\Lambda}(k) \\ &= \frac{1}{\sqrt{\omega(k)}m} \|\psi\| \chi_{\Lambda}(k), \end{aligned} \quad (143)$$

and the estimate

$$(\psi, P^+ H_f P^+ \psi) \leq \frac{4\pi \Lambda^3}{3m^2} \|\psi\|^2 \quad (144)$$

follows. $\square$

References

1. Balinsky, A.A., Evans, W.D.: Stability of one-electron molecules in the Brown-Ravenhall model. Commun. Math. Phys. **202**, 481–500 (1999)
2. Birman, M.S., Koplienko, L.S., Solomyak, M.Z.: Estimates for the spectrum of the difference between fractional powers of two self-adjoint operators. Soviet Mathematics **19**, 1–6 (1975). Translation of Izvestija vyssich.
3. Brown, G., Ravenhall, D.: On the interaction of two electrons. Proc. Roy. Soc. London A **208A**, 552–559 (1951)
4. Bugliaro, L., Fröhlich, J., Graf, G.M.: Stability of quantum electrodynamics with nonrelativistic matter. Phys. Rev. Lett. **77**, 3494–3497 (1996)
5. Daubechies, I., Lieb, E.H.: One-electron relativistic molecules with Coulomb interaction. Commun. Math. Phys. **90**, 497–510 (1983)
6. Evans, W.D., Perry, P., Siedentop, H.: *The spectrum of relativistic one-electron atoms according to Bethe and Salpeter*. Commun. Math. Phys. **178**, 733–746 (1996)
7. Fefferman, C., Fröhlich, J., Graf, G.M.: Stability of nonrelativistic quantum mechanical matter coupled to the (ultraviolet cutoff) radiation field. Proc. Natl. Acad. Sci. USA **93**, 15009–15011 (1996); Stability of ultraviolet cutoff quantum electrodynamics with non-relativistic matter., Commun. Math. Phys. **190**, 309–330 (1997)
8. Griesemer, M., Lieb, E.H., Loss, M.: Ground states in non-relativistic quantum electrodynamics. Invent. Math. **145**, 557–595 (2001)
9. Griesemer, M., Tix, C.: Instability of a pseudo-relativistic model of matter with self-generated magnetic field. J. Math. Phys. **40**, 1780–1791 (1999)
10. Hoever, G., Siedentop, H.: Stability of the Brown-Ravenhall operator. Math. Phys. Electronic Jour. **5**, 1–11 (1999)
11. Kato, T.: *Perturbation Theory for Linear Operators*. Berlin–Heidelberg–New York: Springer Verlag, 1966, p. 307, Remark 5.12
12. Lieb, E.H., Loss, M.: *Analysis*. Providence, RI: Am. Math. Soc. second edition, 2001
13. Lieb, E.H., Loss, M.: *Self-energy of electrons in non-perturbative QED*. In: *Differential Equations and Mathematical Physics, University of Alabama, Birmingham, 1999*, R. Weikard and G. Weinstein, eds. Providence/Cambridge: Amer. Math. Soc./Internat. Press, 2000, pp. 255–269. arXiv math-ph/9908020, mp_arc 99-305
14. Lieb, E.H., Loss, M.: *The Ultraviolet problem in on-relativistic QED*. In preparation
15. Lieb, E.H., Loss, M.: A bound on binding energies and mass renormalization in models of quantum electrodynamics. J. Stat. Phys. **108**, 1057–1069 (2002)
16. Lieb, E.H., Loss, M., Solovej, J.P.: Stability of matter in magnetic fields. Phys. Rev. Lett. **75**, 985–989 (1995)
17. Lieb, E.H., Siedentop, H., Solovej, J.P.: Stability and instability of relativistic electrons in magnetic fields. J. Stat. Phys. **89**, 37–59 (1997)
18. Lieb, E.H., Thirring, W.E.: *Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities*. In E.H. Lieb, B. Simon, and A.S. Wightman, eds., *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*. Princeton, NJ: Princeton University Press, 1976
19. Lieb, E.H., Yau, H.-T.: The stability and instability of relativistic matter. Commun. Math. Phys. **118**, 177–213 (1988)
20. Loss, M.: *Stability of matter in magnetic fields* In: The Proceedings of the XII-th International Congress of Mathematical Physics 1997, De Wit et al eds. Cambridge, MA: International Press, 1999, pp. 98–106
21. Reed, M., Simon, B.: *Methods of Modern Mathematical Physics*. Vol. I, *Functional Analysis*. New York: Academic Press, 1972
22. Reed, M., Simon, B.: *Methods of Modern Mathematical Physics*, Vol. II. *Fourier Analysis, Selfadjointness*. New York: Academic Press, 1975
23. Sucher, J.: *Foundations of the relativistic theory of many-electron atoms*. Phys. Rev. A **22**, 348–362 (1980)
24. Tix, C.: Lower bound for the ground state energy of the no-pair Hamiltonian. Phys. Lett. B **405**, 293–296 (1997)
25. Tix, C.: Strict positivity of a relativistic Hamiltonian due to Brown and Ravenhall. Bull. London Math. Soc. **30**, 283–290 (1998)

Communicated by M. Aizenman

Part VIII

Bosonic Systems

Bosonic Systems

The theme of stability of matter, which is mainly about fermions (electrons) interacting with each other through the electrostatic Coulomb potential, leads, in a natural way, to the study of such particles also interacting with the full, quantized electromagnetic field. This is contained in Part VII.

A related, but different direction of study is the stability and macroscopic properties of matter composed of Bosonic particles (mainly atoms). When such particles interact with short-range forces they behave nicely, but when they interact via Coulomb forces they do not display thermodynamic stability, or, in the relativistic case, not even a finite ground state energy.

Part VIII contains papers on the ground state energies of several kinds of many-boson systems. A comprehensive review of much of the material in this section is *The Quantum-Mechanical Many-Body Problem: The Bose Gas*, #293 in the publication list, and which is reproduced in the *Selecta on Condensed Matter Physics*, edited by B. Nachtergaele, J.P. Solovej and J. Yngvason. That selecta also contains some of the earlier work mentioned below. The review #293 replaces #267, which appeared as VIII.4 in the third edition.

Item VIII.2 *Ground state energy of the low density Bose gas*, with J. Yngvason, proves a conjecture that goes back to Bogolubov's 1947 work: At low particle density ρ the ground state energy per particle, $e_0(\rho)$, for a system of (3-dimensional) bosons with a short-range, two-body potential v of scattering length a is $e_0(\rho) \sim 4\pi(\hbar^2/2m)\rho a$. Actually, the proof in VIII.2 is for nonnegative v (such as a hard core potential) and the question of what happens otherwise is still open. The new conjecture made in the paper is that the formula holds for any short-range v with the property that there are no many-body bound states.

This work was extended in two directions. One direction is to two-dimensions. The relevant formula was proposed only as late as 1971, by Schick: $e_0(\rho) \sim 4\pi(\hbar^2/2m)\rho/|\ln(\rho a^2)|$. This is proved in *The Ground State Energy of a Dilute Two-dimensional Bose Gas*, with J. Yngvason, #258 in the publication list.

The second direction concerns bosons in a trap, as in the experimental situation. To model the trap a one-body confining potential $V(x)$ is added to the Hamiltonian. For low average density, the ground state energy is supposedly described by the Gross-Pitaevskii energy. This is proved in two papers (with J. Yngvason and R. Seiringer): Item VIII.3 *Bosons in a Trap: A Rigorous Derivation of the Gross-Pitaevskii Energy Functional*, which is reproduced here, and *A Rigorous Derivation of the Gross-Pitaevskii Energy Functional for a Two-dimensional Bose Gas*, #263 in the publication list.

Low density gases in traps display interesting features, which include 100% Bose-Einstein condensation and 100% superfluidity. These properties have been demonstrated (in the Gross-Pitaevskii limit) in #273 (with R. Seiringer) and #278 (with R. Seiringer and J. Yngvason) and they are reproduced here as VIII.8 and VIII.9. While the question of Bose-Einstein condensation has now been fully resolved for traps, the existence of Bose-Einstein condensation in the usual thermodynamic limit remains an open problem.

A system of negatively charged bosons with fixed nuclei, as in the stability of matter problem, significantly fails the stability condition. The energy is proportional to $-N^{5/3}$ instead of $-N$ (the $N^{5/3}$ law). This is in VIII.4. If on the other hand, the positive particles have a finite mass and are therefore dynamic, the situation improves somewhat, but not enough. The energy is proportional to $-N^{7/5}$. F. Dyson proved an upper bound of this kind in 1967 and the lower bound was done in 1988 with J. Conlon and H-T. Yau and is in VIII.5. (With a ‘relativistic’ kinetic energy, on the other hand, even this form of stability fails, as shown in V.7. The energy is always $-\infty$ for sufficiently many bosons.)

For charged bosons in a neutralizing background (“jellium”), the “weak-coupling” situation corresponds to *high* density instead of low density. The asymptotic formula was proposed by L. Foldy in 1961 on the basis of Bogolubov’s 1947 theory: $e_0(\rho) \sim C\rho^{1/4}$, with C being a definite constant. This $-\rho^{1/4}$ law was proved in VIII.5, but the correct ‘Foldy constant’ was proved only in 2001 (with J. P. Solovej) in *Ground State Energy of the One-Component Charged Bose Gas*, #265, which is reproduced here as item VIII.6. The importance of verifying the Foldy constant is that it helps justify part of Bogolubov’s approximate theory.

Very recently, the exact constant for the $-N^{7/5}$ energy of two-component charged bosons was proved with J. P. Solovej in VIII.7. The constant is obtained by solving a mean-field equation (which, in turn, involves the ‘Foldy constant’) proposed by Dyson in 1967. Again, this aids our understanding of the Bogolubov theory.

Earlier work on many-boson systems is in publication list #5, #12-16. These results, along with a survey of the state of the field in 1965, are contained in the review article *The Bose fluid*, #18. #12 and #13 *Exact Analysis of an interacting Bose gas* (with W. Liniger) contain an exact solution for one-dimensional bosons with a repulsive delta-function interaction. This is the only soluble model with short-range, two-body potentials (that satisfy the stability condition), and it verifies Bogolubov’s conjecture for the ground state energy in the weak-coupling limit. #14 and #15 (with W. Liniger and A. Sakakura) give a non-rigorous method for calculating the energy for the jellium and short-range models mentioned above, the significant point being that the calculations are done in configuration-space instead of momentum-space, as was, and continues to be, customary.

The one-dimensional solution is now very relevant to experiments because its characteristic properties are seen in elongated traps. The rigorous derivation of one-dimensional behavior from the three-dimensional Schrödinger equation is in #285, 286. Other papers on bosons are discussed in the Condensed Matter selecta.

Ground State Energy of the Low Density Bose Gas

Elliott H. Lieb¹ and Jakob Yngvason²¹Department of Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544²Institut für Theoretische Physik, Universität Wien, Boltzmannngasse 5, A 1090 Vienna, Austria

(Received 27 October 1997)

Now that the properties of low temperature Bose gases at low density, ρ , can be examined experimentally it is appropriate to revisit some of the formulas deduced by many authors four to five decades ago. One of these is that the leading term in the energy/particle is $2\pi\hbar^2\rho a/m$, where a is the scattering length. Owing to the delicate and peculiar nature of bosonic correlations, four decades of research have failed to establish this plausible formula rigorously. The only known lower bound for the energy was found by Dyson in 1957, but it was 14 times too small. The correct bound is proved here. [S0031-9007(98)05619-1]

PACS numbers: 03.75.Fi, 05.30.Jp, 67.40. -w

With the renewed experimental interest in low density, low temperature Bose gases, some of the formulas posited four and five decades ago have been dusted off and re-examined. One of these is the leading term in the ground state energy. In the limit of small particle density, ρ ,

$$e_0(\rho) \approx \mu 4\pi\rho a, \quad (1)$$

where $e_0(\rho)$ is the ground state energy (g.s.e.) per particle in the thermodynamic limit, a is the scattering length (assumed positive) of the two-body potential v for bosons of mass m , and $\mu \equiv \hbar^2/2m$.

Is Eq. (1) correct? In particular, is it true for the hard-sphere gas? While there have been many attempts at a rigorous proof of (1) in the past 40 years, none has been found so far. Our aim here is to supply that proof for finite range, positive potentials. As remarked below, (1) cannot hold unrestrictedly; more than $a > 0$ is needed.

An upper bound for $e_0(\rho)$ agreeing with (1) is not easy to derive, but it was achieved for hard spheres by a variational calculation [1], which can be extended to include general, positive potentials of finite range. What remained unknown was a good lower bound. The only one available is Dyson's [1], and that is about *fourteen times smaller* than (1). In this paper we shall provide a lower bound of the desired form, and thus prove (1). We can also give explicit error bounds for small enough values of the dimensionless parameter $Y \equiv 4\pi\rho a^3/3$:

$$e_0(\rho) \geq \mu 4\pi\rho a(1 - CY^{1/17}) \quad (2)$$

for some fixed C (which is not evaluated explicitly because C and the exponent $1/17$ are only of academic interest). The bound (2) holds for all *non-negative, finite range, spherical, two-body potentials*. A typical experimental value [2] is $Y \approx 10^{-5}$. Dyson's upper bound is $\mu 4\pi\rho a(1 + 2Y^{1/3})(1 - Y^{1/3})^{-2}$.

We conjecture that (1) requires only a positive scattering length and the absence of any many-body, negative energy bound state. If there are such bound states then (1) is certainly wrong, but this obvious caveat does not seem to have been clearly emphasized before. There is

a "nice" potential with positive scattering length, no two-body bound state, but with a three-body bound state [3].

Our method also obviously applies to the positive temperature free energy [because Neumann boundary conditions give an upper bound to the solution to the heat (or Bloch) equation].

We also give some explicit bounds for *finite* systems, which might be useful for experiments with traps, but we concentrate here on the thermodynamic limit for simplicity. For traps with slowly varying confining potentials, V_{ext} , our method will prove that the leading term in the energy is given by the well known local density approximation [4], which minimizes the gaseous energy (1) plus the confining energy, with respect to $\rho(\mathbf{x})$, namely,

$$\mathcal{E}(\rho) \equiv \int [V_{\text{ext}}(\mathbf{x})\rho(\mathbf{x}) + \mu 4\pi\rho a(\mathbf{x})^2] d^3\mathbf{x}$$

is minimized subject to $\int \rho = N = \text{number of particles}$.

The fact that Dyson's lower bound was not improved for four decades, despite many attempts, attests to the fact that bosons are subtle quantum mechanical objects which can have peculiar correlations unknown to fermions. For example, there is the nonthermodynamic $N^{7/5}$ law for the charged Bose gas that was discovered by Dyson [5], confirmed only 20 years later [6], and which defies any simple physical interpretation.

The first understanding of (1) goes back to Bogoliubov [7], who also introduced the notion of "pairing" in Helium (which resurfaced in the BCS theory for fermions). Later, there were several derivations of (1) (and higher order) [8,9]. The method of the pseudopotential, which is an old idea of Fermi's, was closest to the Bogoliubov analysis. The "exact" pseudopotential was constructed in [10], but it did not help to make this appealing idea more rigorous. Most of the derivations were in momentum space, the exception being [9], which works directly in physical space and which can handle both long and short range potentials. See [11] for a review. All these methods rely on special assumptions about the ground state (e.g., selecting special terms in a perturbation expansion, which

likely diverges) and it is important to derive a fundamental result like (1) without extra assumptions.

In all of this earlier work one key fact was not understood, or at least not clearly stated in connection with the derivation of (1). It is that there are two different regimes, even at low density, with very different physics, even though the simple formula (1) seems to depend only on the scattering length. Recall that the (two-body) scattering length is defined, for a spherically symmetric potential, v , by

$$-\mu u_0''(r) + \frac{1}{2} v(r) u_0(r) = 0, \quad (3)$$

with $u_0(0) = 0$, $u_0(r) > 0$ (which is equivalent to the absence of negative energy bound states, and which is true for non-negative v). As $r \rightarrow \infty$, $u(r) \approx r - a$. [Note the $v/2$ and not v in (3) because of the reduced mass.] Thus, a depends on m in a nontrivial way, and there are two extremes:

Potential energy dominated region.—The hard sphere [$v(r) = \infty$ for $r < a$], is the extreme case here; the scattering length is independent of m , and the energy is mostly (entirely) kinetic. We see this from (1) because $-\mu \partial e_0 / \partial m$ is the kinetic energy (Hellmann-Feynman theorem). In this regime the potential is so dominant that it forces the energy to be mostly kinetic. The ground-state (g.s.) wave function is highly correlated.

Kinetic energy dominated region.—The typical case is a very “soft” potential. Then $a \approx (m/\hbar^2) \int_0^\infty v(r) r^2 dr$, which implies, from (1), that e_0 hardly depends on m . Thus, the energy is almost all potential. The g.s. wave function is essentially the noninteracting one in this limit.

In other words, “scattering length” is not a property of v alone, and the low density gas, viewed from the perspective of the bosons, looks quite different in the two regimes. Nevertheless, as (1) says, the energy cannot distinguish the two cases. Whether Bose-Einstein condensation itself can notice the difference remains to be seen. Condensation will not be touched upon here, except to note that so far the *only case with two-body interactions in which Bose-Einstein condensation has been rigorously established is hard core lattice bosons, but only at half filling* [12].

Dyson [1] effectively converted region 1 into region 2. We shall make use of his important idea, which substitutes a very soft potential for the original one (even a hard core) at the price of sacrificing the kinetic energy.

We assume that the N particles are in a $L \times L \times L$ cubic box, Ω . The particle density is then $\rho = NL^{-3}$. It is well known that the energy per particle in the thermodynamic limit, $e_0(\rho)$, does not depend on the details of (reasonable) Ω , so we are free to use a cube and take $N \rightarrow \infty$ through any sequence we please, as far as $e_0(\rho)$ is concerned. We set $N = kM$ with k an integer and M the cube of an integer, because we shall want to divide up Ω into M smaller cubes (called *cells*) of length

$\ell = (k/\rho)^{1/3}$. We will take $M \rightarrow \infty$ with ℓ and $k = \rho \ell^3$ fixed, but large.

The N -body Schrödinger operator is

$$H = -\mu \sum_{i=1}^N \Delta_i + \sum_{1 \leq i < j \leq N} v(\mathbf{x}_i - \mathbf{x}_j). \quad (4)$$

For boundary conditions we impose Neumann (zero derivative) boundary conditions on Ω . It is well known that Neumann boundary conditions give the lowest possible g.s.e. for H , and hence its use is appropriate for a discussion of a lower bound for the g.s.e. Denote this Neumann g.s.e. by $E_0(N, L)$.

Now divide Ω into M cells and impose Neumann conditions on each cell, which, as stated before, lowers the energy further. We also neglect the interaction between particles in different cells; this, too, can only lower the energy because $v \geq 0$.

A lower bound for $E_0(N, L)$ is obtained by distributing the N particles in the M cells and then finding a lower bound for the energy in these cells, which are now independent. We then add these M energies. Finally, we minimize the total energy over *all choices* of the particle number in each cell (subject to the total number being N). Despite the independence of the cells, the latter problem is not easy. In particular, something has to be invoked to make sure that we do not end up with some cells having too large a number of particles and some cells having too few.

With L, N and $M = N/(\rho \ell^3)$ fixed, let $M c_n$, for $n = 0, 1, 2, \dots$ denote the number of cells containing exactly n particles. Then the particle number and cell number constraints are

$$\sum_{n \geq 0} n c_n = k = \rho \ell^3, \quad \sum_{n \geq 0} c_n = 1, \quad (5)$$

and our energy bound is

$$E_0(N, L) \geq M \min_{n \geq 0} c_n E_0(n, \ell), \quad (6)$$

where the minimum is over all $c_n \geq 0$ satisfying (5).

The minimization would be easy if we knew that $E_0(n, \ell)$ (or a good lower bound for it) is convex in n , for then the optimum would be $c_n = \delta_{n,k}$. This convexity is very plausible, but we cannot prove it (except in the thermodynamic limit, where it amounts to thermodynamic stability). What we do know instead is *superadditivity*:

$$E_0(n + n', \ell) \geq E_0(n, \ell) + E_0(n', \ell) \quad (7)$$

for all n, n' , and this turns out to be an adequate substitute for controlling the large n terms in (6).

Equation (7) is an immediate consequence of the positivity of the potential and it is used as follows. Suppose, provisionally, that we have a lower bound of the form

$$E_0(n, \ell) \geq K(\ell) n(n-1), \quad \text{for } 0 \leq n \leq 4k, \quad (8)$$

with $K(\ell)$ independent of n for $0 \leq n \leq 4k$. In fact, we shall later prove that for small enough ρ (and hence small enough k) and suitable ℓ , (8) holds with

$$K(\ell) \geq \mu 4\pi a \ell^{-3}(1 - C'Y^{1/17}), \quad (9)$$

with C' some constant. [However, the analysis we give now, leading to (12), does not depend on this particular form of $K(\ell)$.]

Split the sum in (6) into two pieces: $0 \leq n < 4k$ and $4k \leq n$. Let $t \equiv \sum_{n < 4k} n c_n \leq k$, so that $k - t = \sum_{n \geq 4k} n c_n$. From (8) and Cauchy's inequality (and $\sum_{n < 4k} c_n \leq 1$)

$$\sum_{n < 4k} c_n E_0(n, \ell) \geq K(\ell)t(t-1). \quad (10)$$

On the other hand, if $n \geq 4k$ then, by (7), $E_0(n, \ell) \geq (n/8k)E_0(4k, \ell)$, so

$$\sum_{n \geq 4k} c_n E_0(n, \ell) \geq \frac{k-t}{2} K(\ell)(4k-1).$$

Upon adding (10) and (11) the factor $t(t-1) + (k-t)(2k-1/2)$ is obtained. Although the number t is unknown, we note that this factor is monotone decreasing in t in the interval $0 \leq t \leq k$ [which is where t lies, by (5)]. Thus, we can set $t = k$ and obtain the same bound as if we had convexity, i.e.,

$$E_0(N, L) \geq NK(\ell)(\rho \ell^3 - 1). \quad (12)$$

In summary, if we can show (8) for a box of a fixed size ℓ , for all particle numbers up to $n = 4\rho \ell^3$, then we will have obtained our goal, (2), in the thermodynamic limit provided we can show that the K in (8) satisfies (9) with the constant C' when ℓ is large compared to the mean particle spacing, i.e., $\rho \ell^3 > C''Y^{-1/17}$. Then the C in (2) equals $C' + C''$.

We now focus on a single cell and denote the n coordinates $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ collectively by $\mathbf{X}$. The first step in proving (9) is to replace the total potential, $\sum_{i < j} v(\mathbf{x}_i - \mathbf{x}_j)$, by a lesser quantity, $\mathcal{W}_v(\mathbf{X})$, the nearest-neighbor potential defined by

$$\mathcal{W}_v(\mathbf{X}) \equiv \frac{1}{2} \sum_{i=1}^n v(\mathbf{x}_i - \mathbf{x}_{j(i)}), \quad (13)$$

where $j(i)$ is the nearest-neighbor to particle i in the configuration $\mathbf{X}$; i.e., particle i "feels" only its nearest neighbor. Hence, we replace H by the smaller operator

$$\tilde{H} \equiv \mathcal{T} + \mathcal{W}_v \leq H, \quad (14)$$

where $\mathcal{T} = -\mu \sum_j \Delta_j$ is the kinetic energy in (4). Since $v \geq 0$, the g.s.e. of $\tilde{H}$ satisfies $\tilde{E}_0(n, \ell) \leq E_0(n, \ell)$.

To get into the kinetic energy dominated region, we wish to replace v in (13) by a gentler potential U . To this end we generalize Lemma 1 of [1] and simplify its proof.

LEMMA 1.—Let $v(r) \geq 0$ and $v(r) = 0$ for $r > R_0$. Let $U(r) \geq 0$ be any function satisfying $\int U(r)r^2 dr \leq 1$ and $U(r) = 0$ for $r < R_0$. Let $\mathcal{B} \subset \mathbf{R}^3$ be star-shaped

(convex suffices) with respect to 0. Then, for all functions ϕ ,

$$\int_{\mathcal{B}} \mu |\nabla \phi(\mathbf{x})|^2 + \left[\frac{1}{2} v(r) - \mu a U(r) \right] |\phi(\mathbf{x})|^2 d^3 \mathbf{x} \geq 0. \quad (15)$$

Proof.—Actually, (15) holds with $\mu |\nabla \phi(\mathbf{x})|^2$ replaced by the (smaller) radial kinetic energy, $\mu |\partial \phi(\mathbf{x})/\partial r|^2$, and thus it suffices to prove the analog of (15) for the integral along each radial line, and to assume that $\phi(\mathbf{x}) = u(r)/r$ along this line, with $u(0) = 0$. Let us first prove (15) when U is a delta function at some radius $R \geq R_0$, i.e., $U(r) = R^{-2} \delta(r - R)$. Then, it is enough to show, for all u , that

$$\int_0^R \mu |u'(r) - u(r)/r|^2 + \frac{1}{2} v(r) |u(r)|^2 dr \geq \mu a |u(R)|^2 R^{-2}. \quad (16)$$

If the length of the radial line is less than R then (16) is trivial. Otherwise, normalize u by $u(R) = R - a$, and ask for the minimum of the left side of (16) under the condition that $u(0) = 0$, $u(R) = R - a$. This is a simple problem in the calculus of variations and leads to the scattering length equation (3). If we substitute the solution into (16), integrate by parts, and note that $u_0(r) = r - a$ for $r > R_0$, we find that (16) is true if $a \leq R$, which is true since $u_0 \geq 0$. Finally, by linearity and the fact that $U(r) = \int r^{-2} \delta(r - s) U(s) s^2 ds$, the δ -function case implies the general case. Q.E.D.

We select our U by picking some $R \gg R_0$ and setting

$$U(r) = 3(R^3 - R_0^3)^{-1} \quad \text{for } R_0 < r < R \quad (17)$$

and $U(r) = 0$ otherwise. Later on we shall choose R , and we shall take

$$R_0 \ll R \ll \rho^{-1/3} \ll \ell. \quad (18)$$

By further decomposing a cube into Voronoi cells (which are always convex), Dyson [1] deduces from Lemma 1 that $\tilde{H}$ is bounded below by a nearest-neighbor potential, as in (13), i.e.,

$$H > \tilde{H} > \mu a \mathcal{W}_U(\mathbf{X}), \quad (19)$$

where $\mathcal{W}_U$ is as in (13), with v replaced by U . For the hard core case, Dyson estimates the minimum (over all $\mathbf{X}$) of $\mathcal{W}_U(\mathbf{X})$, for a U similar to (17), and gets a lower bound for all ρ , but 14 times smaller than (1). We follow another route. An important quantity for us will be the average value of $\mathcal{W}_U(\mathbf{X})$ in a cell, denoted by $\langle \mathcal{W}_U \rangle$.

To compute $\langle U(\mathbf{x}_1 - \mathbf{x}_{j(1)}) \rangle$, for example, it is easiest to do the $\mathbf{x}_2, \dots, \mathbf{x}_n$ integrations over the cell first and then the $\mathbf{x}_1$ integration. Provided $\mathbf{x}_1$ is in the smaller cube which is a distance R from the cell boundary [whose volume is $(\ell - 2R)^3$], the probability that $R_0 < |\mathbf{x}_j - \mathbf{x}_1| < R$ is $4\pi(R^3 - R_0^3)/3\ell^3$. Thus, performing the $\mathbf{x}_1$ integration over the smaller cube, and then adding similar

contributions from $U(\mathbf{x}_2 - \mathbf{x}_{j(2)})$, etc., and using (17), we get

$$\langle \mathcal{W}_U \rangle \geq \frac{3n(\ell - 2R)^3}{(R^3 - R_0^3)\ell^3} [1 - (1 - Q)^{n-1}] \quad (20)$$

$$\geq \frac{4\pi}{\ell^3} n(n-1) \left(1 - \frac{2R}{\ell}\right)^3 \frac{1}{1 + Q(n-1)}. \quad (21)$$

with

$$Q = 4\pi(R^3 - R_0^3)/3\ell^3 \ll 1. \quad (22)$$

In (21) we used $[1 - x]^{n-1} \leq [1 + (n-1)x]^{-1}$ for $0 \leq x \leq 1$. Note that (21) is of the form (8).

By similar reasoning, we obtain the upper bound

$$\langle \mathcal{W}_U \rangle \leq \frac{3n}{R^3 - R_0^3} [1 - (1 - Q)^{n-1}] \quad (23)$$

$$\leq \frac{4\pi}{\ell^3} n(n-1). \quad (24)$$

Since $U(r)^2 = 4\pi(Q\ell^3)^{-1}U(r)$, we also obtain

$$\langle \mathcal{W}_U^2 \rangle \leq 4\pi n(Q\ell^3)^{-1} \langle \mathcal{W}_U \rangle. \quad (25)$$

We can now use Lemma 1 and these averages to obtain (8) and (9). Instead of using (19) alone, we pick some $0 < \varepsilon \ll 1$ and, borrowing a bit of kinetic energy, define

$$\hat{H} \equiv \varepsilon \mathcal{T} + (1 - \varepsilon)\mu a \mathcal{W}_U(\mathbf{X}). \quad (26)$$

By Lemma 1 and $v \geq 0$, we have

$$H > \tilde{H} > \hat{H}. \quad (27)$$

We shall derive (8) and (9) from a lower bound to $\hat{H}$.

Although ε is small, we regard $H_0 \equiv \varepsilon \mathcal{T}$ as our unperturbed Hamiltonian and $V \equiv (1 - \varepsilon)\mu a \mathcal{W}_U(\mathbf{X})$ as a perturbation of H_0 . The ground state wave function for H_0 is $\Psi_0(\mathbf{X}) = \ell^{-3n/2}$ and $H_0\Psi_0 = \lambda_0\Psi_0 = 0$ (Neumann conditions). The second eigenvalue of H_0 is $\lambda_1 = \varepsilon\mu\pi/\ell^2$. Note that the ground state expectation, $\langle \Psi_0 | \mathcal{W}_U | \Psi_0 \rangle$, is precisely the average $\langle \mathcal{W}_U \rangle$ mentioned in (20)–(25).

Temple's inequality [13] states that when a perturbation V is non-negative (as here) and when $\lambda_1 - \lambda_0 \geq \langle \Psi_0 | V | \Psi_0 \rangle$ then the g.s.e., E_0 , of the perturbed Hamiltonian $H = H_0 + V$ satisfies

$$E_0 \geq \lambda_0 + \langle \Psi_0 | V | \Psi_0 \rangle - \frac{\langle \Psi_0 | V^2 | \Psi_0 \rangle - \langle \Psi_0 | V | \Psi_0 \rangle^2}{\lambda_1 - \lambda_0 - \langle \Psi_0 | V | \Psi_0 \rangle}. \quad (28)$$

We apply this to our case with $\lambda_1 - \lambda_0 = \varepsilon\mu\pi/\ell^2$ and $V = (1 - \varepsilon)\mu a \mathcal{W}_U$. We neglect the (positive) term $\langle \Psi_0 | V | \Psi_0 \rangle^2$ in (28) and we use (21), (24), and (25). We

also use $1 - \varepsilon < 1$ in two appropriate places and find

$$\frac{E_0(n, \ell)}{\mu a \langle \mathcal{W}_U \rangle} \geq (1 - \varepsilon) \left(1 - \frac{4\pi a n}{Q\ell} \frac{1}{\varepsilon\pi - a\ell^2 \langle \mathcal{W}_U \rangle} \right). \quad (29)$$

Apart from some higher order errors, (29) is just what we need in (8) and (9). Let us denote the order of the main error by Y^α , and we would like to show that $\alpha = 1/17$ suffices. The errors are the following:

From the $(1 - \varepsilon)$ factor, we need $\varepsilon \leq O(Y^\alpha)$.

From the $Q(n+1)$ error in (21) we need $Q\rho\ell^3 \leq O(Y^\alpha)$.

From the R/ℓ error in (21) we need $R/\ell \leq O(Y^\alpha)$.

From (29) we need $a\ell^2 \langle \mathcal{W}_U \rangle / \varepsilon \leq O(Y^\alpha)$ and $\rho\ell^5 a / R^3 \varepsilon \leq O(Y^\alpha)$.

All these desiderata can be met with $\varepsilon = Y^\alpha$, $R/\ell = Y^\alpha$, $Q = O(Y^\alpha)$, $\rho R^3 = Y^{2\alpha}$, and $\alpha = 1/17$ —as claimed.

The partial support of U.S. National Science Foundation Grant No. PHY95-13072A01 (EHL) and the Adalsteinn Kristjansson Foundation of the University of Iceland (JY) is gratefully acknowledged.

- [1] F. J. Dyson, Phys. Rev. **106**, 20 (1957).
- [2] W. Ketterle and N. J. van Druten, in *Advances in Atomic, Molecular and Optical Physics*, edited by B. Bederson and H. Walther (Academic Press, New York, 1996), Vol. 37, p. 181.
- [3] B. Baumgartner, J. Phys. A **30**, L741 (1997).
- [4] J. Oliva, Phys. Rev. B **39**, 4197 (1989).
- [5] F. J. Dyson, J. Math. Phys. (N.Y.) **8**, 1538 (1967).
- [6] J. G. Conlon, E. H. Lieb, and H.-T. Yau, Commun. Math. Phys. **116**, 417 (1988).
- [7] N. N. Bogoliubov, J. Phys. (U.S.S.R.) **11**, 23 (1947); N. N. Bogoliubov and D. N. Zubarev, Sov. Phys. JETP **1**, 83 (1955).
- [8] K. Huang and C. N. Yang, Phys. Rev. **105**, 767–775 (1957); T. D. Lee, K. Huang, and C. N. Yang, Phys. Rev. **106**, 1135–1145 (1957); K. A. Brueckner and K. Sawada, Phys. Rev. **106**, 1117–1127 (1957); **106**, 1128–1135 (1957); S. T. Beliaev, Sov. Phys. JETP **7**, 299–307 (1958); T. T. Wu, Phys. Rev. **115**, 1390 (1959); N. Hugenholtz and D. Pines, Phys. Rev. **116**, 489 (1959); M. Girardeau and R. Arnowitt, Phys. Rev. **113**, 755 (1959); T. D. Lee and C. N. Yang, Phys. Rev. **117**, 12 (1960).
- [9] E. H. Lieb, Phys. Rev. **130**, 2518 (1963); E. H. Lieb and A. Y. Sakakura, Phys. Rev. **133**, A899 (1964); E. H. Lieb and W. Liniger, Phys. Rev. **134**, A312 (1964).
- [10] E. H. Lieb, Proc. Nat. Acad. Sci. U.S.A. **46**, 1000 (1960).
- [11] E. H. Lieb, *The Bose Fluid*, in Lecture Notes in Theoretical Physics VIII, edited by W. E. Brittin (University of Colorado Press, Boulder, Colorado, 1964), p. 175.
- [12] E. H. Lieb, T. Kennedy, and S. Shastry, J. Stat. Phys. **53**, 1019 (1988).
- [13] G. Temple, Proc. R. Soc. London A **119**, 276 (1928).

Bosons in a trap: A rigorous derivation of the Gross-Pitaevskii energy functional

Elliott H. Lieb

Department of Physics and Department of Mathematics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544

Robert Seiringer and Jakob Yngvason

Institut für Theoretische Physik, Universität Wien, Boltzmannsgasse 5, A 1090 Vienna, Austria

(Received 31 August 1999; published 6 March 2000)

The ground-state properties of interacting Bose gases in external potentials, as considered in recent experiments, are usually described by means of the Gross-Pitaevskii energy functional. We present here a rigorous proof of the asymptotic exactness of this approximation for the ground-state energy and particle density of a dilute Bose gas with a positive interaction.

PACS number(s): 03.75.Fi, 05.30.Jp, 67.40.Db, 71.35.Lk

I. INTRODUCTION

Recent experimental breakthroughs in the treatment of dilute Bose gases have renewed interest in formulas for the ground state and its energy derived many decades ago. One of these is the Gross-Pitaevskii (GP) formula for the energy in a trap [1–3], such as is used in the actual experiments. We refer to [4] for an up-to-date review of this approximation and its applications. One of the inputs needed for its justification is the ground-state energy per unit volume of a dilute, thermodynamically infinite, homogeneous gas. This latter quantity has been known for many years, but it was only very recently that it was derived rigorously [5] for suitable interparticle potentials. Consequently, it is appropriate now to use this result to go one step further and derive the GP formula rigorously.

The starting point for our investigation is the Hamiltonian for N identical bosons

$$H^{(N)} = \sum_{i=1}^N [-\nabla_i^2 + V(\mathbf{x}_i)] + \sum_{i < j} v(|\mathbf{x}_i - \mathbf{x}_j|) \quad (1.1)$$

acting on totally symmetric, square integrable wave functions of $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ with $\mathbf{x}_i \in \mathbb{R}^3$. Units have here been chosen so that $\hbar = 2m = 1$, where m is the mass. We consider external potentials V that are measurable and locally bounded and tend to infinity for $|\mathbf{x}| \rightarrow \infty$ in the sense that $\inf_{|\mathbf{x}| \geq R} V(\mathbf{x}) \rightarrow \infty$ for $R \rightarrow \infty$. The potential is then bounded below and for convenience we assume that its minimum value is zero. The ground state of $-\nabla^2 + V(\mathbf{x})$ provides a natural energy unit, $\hbar\omega$, and the corresponding length unit, $\sqrt{\hbar/m\omega}$, describes the extension of the potential. We shall measure all energies and lengths in these units. In the available experiments V is typically $\sim |\mathbf{x}|^2$ and $\sqrt{\hbar/m\omega}$ of the order 10^{-6} m.

The particle interaction v is assumed to be positive, spherically symmetric, and decay faster than $|\mathbf{x}|^{-3}$ at infinity. In particular, the scattering length, denoted by a , should be finite. We recall that the (two-body) scattering length is defined by means of the solution $u(r)$ of the zero-energy scattering equation

$$-u''(r) + \frac{1}{2}v(r)u(r) = 0 \quad (1.2)$$

with $u(0)=0$; by definition, $a = \lim_{r \rightarrow \infty} [r - u(r)/u'(r)]$.

Let $v_1(r)$ be a fixed potential with scattering length a_1 . Then $v(r) = (a_1/a)^2 v_1(a_1 r/a)$ has scattering length a . In the following we regard v_1 as fixed, but vary a (in fact, we shall take $a = a_1/N$). The ground-state energy E^{QM} of Eq. (1.1) depends on the potentials V and v , besides N , but with V fixed and $v(r) = (a_1/a)^2 v_1(a_1 r/a)$, the notation $E^{\text{QM}}(N, a)$ is justified. The corresponding eigenfunction will be denoted $\Psi_0^{(N)}$. It is unique up to a phase that can be chosen such that the wave function is strictly positive where the interaction is finite [6]. The particle density is defined by

$$\rho_{N,a}^{\text{QM}}(\mathbf{x}) = N \int_{\mathbb{R}^{3(N-1)}} |\Psi_0^{(N)}(\mathbf{x}, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 d\mathbf{x}_2 \cdots d\mathbf{x}_N. \quad (1.3)$$

The Gross-Pitaevskii (GP) energy functional is defined as

$$\mathcal{E}^{\text{GP}}[\Phi] = \int [|\nabla \Phi(\mathbf{x})|^2 + V(\mathbf{x})|\Phi(\mathbf{x})|^2 + 4\pi a |\Phi(\mathbf{x})|^4] d\mathbf{x} \quad (1.4)$$

where Φ is a function on $\mathbb{R}^3$. For a given N the corresponding GP energy, denoted $E^{\text{GP}}(N, a)$, is defined as the infimum of $\mathcal{E}[\Phi]$ under the normalization condition

$$\int_{\mathbb{R}^3} |\Phi(\mathbf{x})|^2 d\mathbf{x} = N. \quad (1.5)$$

It has the simple scaling property

$$E^{\text{GP}}(N, a) = N E^{\text{GP}}(1, Na). \quad (1.6)$$

What Eq. (1.6) shows is that the GP functional (1.4) together with the normalization condition (1.5) has one characteristic parameter, namely, Na . (Recall that lengths are measured in the unit $\sqrt{\hbar/m\omega}$ associated with V so a is dimensionless.) Thus, if we want to investigate the nontrivial aspects of GP theory we have to consider a limit in which $N \rightarrow \infty$ with Na fixed. This explains the seemingly peculiar limit in Theorems I.1 and I.2. As $Na \rightarrow \infty$ the GP energy functional simplifies, since the gradient term becomes small compared to the other terms, and the so called Thomas-Fermi limit described in Theorem II.2 results. In some typical experiments a is about

10^{-3} , while N varies from 10^3 to 10^7 . Thus a_1 in Theorems I.1 and I.2 varies from 1 to about 10^4 .

In the next section it will be shown that the infimum of the energy functional (1.4), under the subsidiary condition (1.5), is obtained for a unique, strictly positive function, denoted Φ^{GP} . The GP density is, by definition,

$$\rho_{N,a}^{\text{GP}}(\mathbf{x}) = \Phi^{\text{GP}}(\mathbf{x})^2. \quad (1.7)$$

It satisfies

$$\rho_{N,a}^{\text{GP}}(\mathbf{x}) = N \rho_{1,N,a}^{\text{GP}}(\mathbf{x}). \quad (1.8)$$

The main result of this paper concerns the behavior of the quantum mechanical ground-state energy $E^{\text{QM}}(N,a)$ when N is large, but a is small, so that Na is $O(1)$. It is important to note that although the density tends to infinity for $N \rightarrow \infty$ [by Eq. (1.8)] we are still concerned with *dilute* systems in the sense that $a^3 \bar{\rho} \ll 1$, where

$$\bar{\rho} = \frac{1}{N} \int \rho_{N,a}^{\text{GP}}(\mathbf{x})^2 d\mathbf{x} \quad (1.9)$$

is the *mean GP density*. [Note the exponent 2 in Eq. (1.9).] In fact, since $a \sim N^{-1}$, $a^3 \bar{\rho} \sim N^{-2}$.

The precise statement of the limit theorem for the energy is as follows.

Theorem I.1 (The GP energy is the dilute limit of the QM energy). For every fixed a_1

$$\lim_{N \rightarrow \infty} \frac{E^{\text{QM}}(N, a_1/N)}{N} = E^{\text{GP}}(1, a_1) \quad (1.10)$$

and the convergence is uniform on bounded intervals of a_1 .

While we do not prove anything about Bose-Einstein condensation, which necessarily involves the full one-body density matrix $\rho^{(1)}(\mathbf{x}, \mathbf{x}')$, we can make an assertion about the diagonal part of the density matrix, $\rho^{\text{QM}}(\mathbf{x}) = \rho^{(1)}(\mathbf{x}, \mathbf{x})$:

Theorem I.2 (The GP density is the dilute limit of the QM density). For every fixed a_1

$$\lim_{N \rightarrow \infty} \frac{1}{N} \rho_{N,a_1/N}^{\text{QM}}(\mathbf{x}) = \rho_{1,a_1}^{\text{GP}}(\mathbf{x}) \quad (1.11)$$

in the sense of weak convergence in L^1 .

For the proof of Theorem I.1 we establish upper and lower bounds on $E^{\text{QM}}(N,a)$ in terms of $E^{\text{GP}}(N,a)$ with controlled errors. Theorem I.2 follows from Theorem I.1 by variation of the external potential. The upper bound is obtained in Sec. III by a variational calculation that generalizes the upper bound of Dyson [7] for a homogeneous gas of hard spheres. We also derive an upper bound on the chemical potential, i.e., the energy increase when one particle is added to the system. This upper bound is used in the proof of the lower bound of the energy in Sec. IV. The main ingredient for the lower bound, however, is the bound for the homogeneous case established in [5]. In addition, some basic properties of the minimizer of the GP functional are used in the proof and we consider them next.

II. THE GROSS-PITAEVSKII ENERGY FUNCTIONAL

The GP functional is defined by Eq. (1.4) for $\Phi \in \mathcal{D}$ with

$$\mathcal{D} = \{\Phi: \nabla \Phi \in L^2(\mathbb{R}^3), V|\Phi|^2 \in L^1(\mathbb{R}^3),$$

$$\Phi \in L^4(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)\}, \quad (2.1)$$

where $f \in L^p(\mathbb{R}^n)$ means $\int_{\mathbb{R}^n} |f(\mathbf{x})|^p d\mathbf{x} < \infty$. The corresponding GP energy is given by

$$E^{\text{GP}}(N,a) = \inf\{E^{\text{GP}}[\Phi]: \Phi \in \mathcal{D}_N\} \quad (2.2)$$

with

$$\mathcal{D}_N = \mathcal{D} \cap \left\{ \Phi: \int |\Phi(\mathbf{x})|^2 d\mathbf{x} = N \right\}. \quad (2.3)$$

The basic facts about the GP functional are summarized in the following theorem.

Theorem II.1 (Existence and properties of a minimizer). The infimum in Eq. (2.2) is a minimum, i.e., there is a $\Phi^{\text{GP}} \in \mathcal{D}_N$ such that $E^{\text{GP}}(N,a) = E^{\text{GP}}[\Phi^{\text{GP}}]$. This Φ^{GP} is unique up to a phase factor, which can be chosen so that Φ^{GP} is strictly positive. Φ^{GP} is at least once continuously differentiable, and if V is C^∞ then also Φ^{GP} is C^∞ . The energy $E^{\text{GP}}(N,a)$ is continuously differentiable in a and hence [by Eq. (1.6)] also in N . The minimizer Φ^{GP} solves the Gross-Pitaevskii equation

$$-\nabla^2 \Phi + V\Phi + 8\pi a |\Phi|^2 \Phi = \mu \Phi \quad (2.4)$$

(in the sense of distributions) with

$$\mu = dE^{\text{GP}}(N,a)/dN = E^{\text{GP}}(N,a)/N + 4\pi a \bar{\rho}. \quad (2.5)$$

Here $\bar{\rho}$ is the mean density (1.9).

The GP energy functional is mathematically quite similar to the energy functional of Thomas-Fermi-von Weizsäcker theory and Theorem II.1 can be proved by the methods of Sec. VII in [8]. For completeness, the proof is given in Appendix A. With additional properties of V one can draw further conclusions about Φ^{GP} .

Proposition II.1 (Symmetry and monotonicity). If V is spherically symmetric and monotone increasing, then Φ^{GP} is spherically symmetric and monotone decreasing.

Proof. Let Φ^* be the symmetric-decreasing rearrangement of Φ^{GP} (see [9]). Then $E^{\text{GP}}[\Phi^*] \leq E^{\text{GP}}[\Phi^{\text{GP}}]$. $\square$

Proposition II.2 (Log concavity). If V is convex, then Φ^{GP} is log concave, i.e., $\Phi^{\text{GP}}(\mathbf{x}) \wedge \Phi^{\text{GP}}(\mathbf{y})^{(1-\lambda)} \leq \Phi^{\text{GP}}(\lambda \mathbf{x} + (1-\lambda)\mathbf{y})$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, $\lambda \in (0,1)$.

Proof. Using the Trotter product formula it suffices to show that the solutions $u(t, \mathbf{x})$ of the equations

$$\frac{\partial u}{\partial t} - \nabla^2 u = 0, \quad \frac{\partial u}{\partial t} + Vu = 0, \quad \frac{\partial u}{\partial t} + 8\pi a u^3 = \mu u$$

are log concave, if $u(0, \mathbf{x})$ is log concave. The first follows from the fact that the convolution of two log concave functions is log concave, the second follows easily from convexity of V , and the third is shown in [10]. $\square$

The GP theory has a well defined limit if $Na \rightarrow \infty$. It is sometimes referred to as the "Thomas-Fermi limit" of GP theory because the gradient term vanishes in this limit. For simplicity we restrict ourselves to *homogeneous external potentials* V , i.e.,

$$V(\lambda \mathbf{x}) = \lambda^s V(\mathbf{x}) \quad (2.6)$$

for some $s > 0$.

Theorem II.2 (Large Na limit). *Let V be homogeneous of order s and let $\mathcal{F}$ be the functional*

$$\mathcal{F}[\rho] = \int_{\mathbb{R}^3} [V(\mathbf{x})\rho(\mathbf{x}) + 4\pi a \rho(\mathbf{x})^2] d\mathbf{x} \quad (2.7)$$

with $\rho(\mathbf{x}) \geq 0$, $\mathbf{x} \in \mathbb{R}^3$. Let $F(N, a)$ be the infimum of $\mathcal{F}$ under the condition $\int \rho = N$. By scaling, $F(N, a) = NF(1, Na)$ and $F(1, Na) = (Na)^{s/(s+3)} F(1, 1)$. In the limit $Na \rightarrow \infty$ we have

$$\lim_{Na \rightarrow \infty} \frac{E^{GP}(1, Na)}{(Na)^{s/(s+3)}} = F(1, 1). \quad (2.8)$$

The minimizing density of $\mathcal{F}$ under the condition $\int \rho = 1$ and with $a = 1$ is given by

$$\rho_{1,1}^F(\mathbf{x}) = (8\pi)^{-1} [\tilde{\mu} - V(\mathbf{x})]_+ \quad (2.9)$$

with $\tilde{\mu} = F(1, 1) + 4\pi \int (\rho_{1,1}^F)^2$, and $[t]_+ = t$ for $t > 0$ and 0 otherwise. Moreover,

$$\lim_{Na \rightarrow \infty} \rho_{1,Na}^{GP}(\mathbf{x}) = \rho_{1,1}^F(\mathbf{x}) \quad (2.10)$$

strongly in $L^2(\mathbb{R}^3)$.

Proof. Since $\mathcal{E}^{GP}[\sqrt{\rho}] \geq \mathcal{F}[\rho]$ it is clear that $E^{GP}(1, Na) \geq F(1, Na)$. For the converse we write ρ in the form $\rho(\mathbf{x}) = (Na)^{-3/(s+3)} \tilde{\rho}((Na)^{-1/(s+3)} \mathbf{x})$ and obtain

$$\begin{aligned} \mathcal{E}^{GP}[\sqrt{\rho}] &= (Na)^{s/(s+3)} \int [(Na)^{-(s+2)/(s+3)} |\nabla \sqrt{\tilde{\rho}}|^2 + V\tilde{\rho} \\ &\quad + 4\pi \tilde{\rho}^2] d\mathbf{x}, \end{aligned}$$

$$\mathcal{F}[\rho] = (Na)^{s/(s+3)} \int (V\tilde{\rho} + 4\pi \tilde{\rho}^2) d\mathbf{x}.$$

In particular, $F(1, Na) = (Na)^{s/(s+3)} F(1, 1)$, and with $\tilde{\rho} = \rho_{1,1}^F$ we obtain

$$E^{GP}(1, Na) \leq F(1, Na) + (Na)^{-2/(s+3)} \int |\nabla \sqrt{\rho_{1,1}^F}|^2.$$

(After a slight regularization, we may assume that $\int |\nabla \sqrt{\rho_{1,1}^F}|^2 < \infty$.) In the limit $Na \rightarrow \infty$ the gradient term vanishes, and thus the limit of the energies is proved. Now

$$\begin{aligned} \lim_{Na \rightarrow \infty} \frac{\mathcal{E}^{GP}[\sqrt{\rho_{1,Na}^{GP}}]}{(Na)^{s/(s+3)}} &= \lim_{Na \rightarrow \infty} \left(\frac{\mathcal{F}[\rho_{1,Na}^{GP}]}{(Na)^{s/(s+3)}} \right. \\ &\quad \left. + (Na)^{-(s+2)/(s+3)} \int |\nabla \sqrt{\rho_{1,Na}^{GP}}|^2 \right) \\ &= F(1, 1). \end{aligned}$$

Since $F(1, 1)$ is the minimum of $\mathcal{F}(Na)^{s/(s+3)}$, the second term vanishes for $Na \rightarrow \infty$, and it follows that $\rho_{1,Na}^{GP}$ is a minimizing sequence for $\int (V\rho + 4\pi\rho^2)$. Since both terms in the functional are non-negative, they must converge individually; in particular, $\|\rho_{1,Na}^{GP}\|_2$ converges to $\|\rho_{1,1}^F\|_2$. On the other hand $\rho_{1,Na}^{GP}$ converges weakly to $\rho_{1,1}^F$ by uniqueness of the minimizer. Together with the convergence of the norms this implies strong convergence.

The solution of the variational equation for $\rho_{1,1}^F$ is simply $\rho_{1,1}^F = (8\pi)^{-1} [\tilde{\mu} - V]_+$ with $\tilde{\mu}$ given by $\tilde{\mu} = F(1, 1) + 4\pi \int (\rho_{1,1}^F)^2$. $\square$

Lemma II.1 (Virial theorem). *When V is homogeneous of order s , as in Eq. (2.6), the minimizer of the GP functional satisfies*

$$\begin{aligned} \frac{2}{3} \int |\nabla \Phi^{GP}(\mathbf{x})|^2 d\mathbf{x} - \frac{s}{3} \int \Phi^{GP}(\mathbf{x})^2 V(\mathbf{x}) d\mathbf{x} \\ + 4\pi a \int \Phi^{GP}(\mathbf{x})^4 d\mathbf{x} = 0. \end{aligned} \quad (2.11)$$

Proof. Define Φ_k by

$$\Phi_k(\mathbf{x}) = k^{1/2} \Phi^{GP}(k^{1/3} \mathbf{x}).$$

Because Φ^{GP} is the minimizer of $\mathcal{E}^{GP}[\Phi]$, it must be true that

$$\left. \frac{\partial}{\partial k} \mathcal{E}^{GP}[\Phi_k] \right|_{k=1} = 0.$$

This leads to the virial theorem (2.11). $\square$

In the proof of the lower bound we shall also consider the GP energy functional in a finite box. For $R > 0$ we denote by Λ_R a cube centered at the origin, with side length $2R$. The energy functional $\mathcal{E}_R^{GP}$ in the box is simply Eq. (1.4) with the integration reduced to Λ_R . The corresponding minimizer, denoted by ρ_R^{GP} , satisfies Neumann conditions at the boundary of Λ_R , and is strictly positive. Analogously to Eqs. (1.7), (1.9), and (2.5) we define ρ_R^{GP} , $\bar{\rho}_R$, and μ_R . The corresponding energy will be denoted by $E_R^{GP}(N, a)$. Then we have the following:

Lemma II.2 (GP energy in a box).

$$\lim_{R \rightarrow \infty} E_R^{GP}(N, a) = E^{GP}(N, a). \quad (2.12)$$

Proof. Using $N^{1/2} \Phi^{GP} \chi_R / \|\Phi^{GP} \chi_R\|_2$ as a test function for $\mathcal{E}_R^{GP}$, where χ_R denotes the characteristic function of Λ_R , we immediately get

$$\lim_{R \rightarrow \infty} E_R^{GP}(N, a) \leq E^{GP}(N, a). \quad (2.13)$$

Let Θ_R be a C^∞ function on $\mathbb{R}^3$, with $\Theta_R=0$ outside Λ_R , and $\Theta_R=1$ inside Λ_{R-1} . We use $N^{1/2}\Phi_R^{\text{GP}}\Theta_R/\|\Phi_R^{\text{GP}}\Theta_R\|_2$ as a test function for $\mathcal{E}^{\text{GP}}$. Since $\nabla\Theta_R$ is bounded and

$$\lim_{R \rightarrow \infty} \int_{\Lambda_R \setminus \Lambda_{R-1}} |\Phi_R^{\text{GP}}|^2 = 0$$

[because V tends to infinity and $\mathcal{E}_R^{\text{GP}}[\Phi_R^{\text{GP}}]$ is bounded by Eq. (2.13)], we have

$$E^{\text{GP}}(N, a) \leq \liminf_{R \rightarrow \infty} E_R^{\text{GP}}(N, a).$$

This completes the proof of Eq. (2.12). $\square$

III. UPPER BOUNDS

A. Upper bound for the quantum mechanical energy

It will now be shown that for all N and small values of $a\bar{\rho}^{1/3}$ [with $\bar{\rho} = \int \rho^{\text{GP}}(\mathbf{x})^2 d\mathbf{x}/N$, cf. Eq. (1.9)]

$$E^{\text{QM}}(N, a) \leq E^{\text{GP}}(N, a) [1 + O(a\bar{\rho}^{1/3})]. \quad (3.1)$$

This upper bound, which holds for all positive, spherically symmetric v with finite scattering length, is derived by means of the variational principle. We generalize a method of Dyson [7], who proved an upper bound for the homogeneous Bose gas with hard-sphere interaction. Consider as a trial function

$$\Psi = F(\mathbf{x}_1, \dots, \mathbf{x}_N) G(\mathbf{x}_1, \dots, \mathbf{x}_N) \quad (3.2)$$

with

$$F(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{i=1}^N F_i(\mathbf{x}_1, \dots, \mathbf{x}_i),$$

$$G(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{i=1}^N g(\mathbf{x}_i) \quad (3.3)$$

where

$$F_i(\mathbf{x}_1, \dots, \mathbf{x}_i) = f(t_i),$$

$$t_i = \min\{|\mathbf{x}_i - \mathbf{x}_j|, j = 1, \dots, i-1\}, \quad (3.4)$$

with a function f satisfying

$$0 \leq f \leq 1, f' \geq 0, \quad (3.5)$$

and

$$g(\mathbf{x}) = \Phi^{\text{GP}}(\mathbf{x}) / \|\Phi^{\text{GP}}\|_\infty. \quad (3.6)$$

The function f will be specified later. This trial function is not symmetric in the particle coordinates, but the expectation value $\langle \Psi | H^{(N)} | \Psi \rangle / \langle \Psi | \Psi \rangle$ is still an upper bound to the bosonic ground-state energy because the Hamiltonian is symmetric and its ground-state wave function is positive. Hence the bosonic ground-state energy is equal to the *absolute* ground-state energy [7,11].

The physical meaning of the trial function can be understood as follows: The G part describes independent particles, each with the GP wave function. The F part means that the particles are inserted into the system one at a time, taking into account the particles previously inserted, but without adjusting their wave function (cf. [7]). Although a wave function of this form cannot describe all correlations present in the true ground state, it captures the leading term in the energy for dilute systems.

For the computation of the kinetic energy we use

$$\int_{\mathbb{R}^{3N}} \Psi \nabla_k^2 \Psi = \int_{\mathbb{R}^{3N}} (G \nabla_k^2 G) F^2 - \int_{\mathbb{R}^{3N}} G^2 |\nabla_k F|^2, \quad (3.7)$$

where ∇_k denotes the gradient with respect to $\mathbf{x}_k$, $k = 1, \dots, N$. We write

$$\epsilon_{ik} = \begin{cases} 1 & \text{for } i=k \\ -1 & \text{for } t_i = |\mathbf{x}_i - \mathbf{x}_k| \\ 0 & \text{otherwise.} \end{cases} \quad (3.8)$$

Let n_i be the unit vector in the direction of $\mathbf{x}_i - \mathbf{x}_{j(i)}$, when $\mathbf{x}_{j(i)}$ is the nearest to $\mathbf{x}_i$ of the points $(\mathbf{x}_1, \dots, \mathbf{x}_{i-1})$. [Note that $j(i)$ really depends on all the points $\mathbf{x}_1, \dots, \mathbf{x}_i$ and not just on the index i . Except for a set of zero measure, $j(i)$ is unique.] Then

$$G \nabla_k F = \sum_i \Psi F_i^{-1} \epsilon_{ik} n_i f'(t_i), \quad (3.9)$$

and after summation over k

$$\begin{aligned} \sum_k G^2 |\nabla_k F|^2 &= |\Psi|^2 \sum_{i,j,k} \epsilon_{ik} \epsilon_{jk} (n_i \cdot n_j) F_i^{-1} F_j^{-1} f'(t_i) f'(t_j) \\ &\leq 2 |\Psi|^2 \sum_i F_i^{-2} f'(t_i)^2 + 2 |\Psi|^2 \\ &\quad \times \sum_{k \leq i < j} |\epsilon_{ik} \epsilon_{jk}| F_i^{-1} F_j^{-1} f'(t_i) f'(t_j). \end{aligned} \quad (3.10)$$

The expectation value can thus be bounded as follows:

$$\begin{aligned} \frac{\langle \Psi | H^{(N)} | \Psi \rangle}{\langle \Psi | \Psi \rangle} &\leq 2 \sum_{i=1}^N \frac{\int |\Psi|^2 F_i^{-2} f'(t_i)^2}{\int |\Psi|^2} \\ &\quad + \sum_{j < i} \frac{\int |\Psi|^2 v(|\mathbf{x}_i - \mathbf{x}_j|)}{\int |\Psi|^2} \\ &\quad + 2 \sum_{k \leq i < j} \frac{\int |\Psi|^2 |\epsilon_{ik} \epsilon_{jk}| F_i^{-1} F_j^{-1} f'(t_i) f'(t_j)}{\int |\Psi|^2} \\ &\quad + \sum_{i=1}^N \frac{\int |\Psi|^2 [-g(\mathbf{x}_i) \nabla_i^2 g(\mathbf{x}_i) + V(\mathbf{x}_i)]}{\int |\Psi|^2}. \end{aligned} \quad (3.11)$$

For $i < p$, let $F_{p,i}$ be the value that F_p would take if the point $\mathbf{x}_i$ were omitted, i.e.,

$$F_{p,i} = f(|\mathbf{x}_p - \mathbf{x}_{k(p)}|), \quad (3.12)$$

where $\mathbf{x}_{k(p)}$ is the nearest to $\mathbf{x}_p$ of the points $(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_{p-1})$. The reason for introducing these functions is that one wants to decouple the integration over $\mathbf{x}_i$ from the integrations over the other variables. (Note that $F_{p,i}$ is independent of $\mathbf{x}_i$.) Analogously, one defines $F_{p,i,j}$ by omitting $\mathbf{x}_i$ and $\mathbf{x}_j$. This simultaneously decouples $\mathbf{x}_i$ and $\mathbf{x}_j$ from the other variables. The functions F_i occur both in the numerator and the denominator so one needs estimates from below and above. Since

$$F_p = \min\{F_{p,i,j}, f(|\mathbf{x}_p - \mathbf{x}_j|), f(|\mathbf{x}_p - \mathbf{x}_i|)\}, \quad (3.13)$$

one has (recall that $f \leq 1$)

$$F_{p,i,j}^2 f(|\mathbf{x}_p - \mathbf{x}_i|)^2 f(|\mathbf{x}_p - \mathbf{x}_j|)^2 \leq F_p^2 \leq F_{p,i,j}^2. \quad (3.14)$$

Hence, with $j < i$,

$$F_{j+1}^2 \cdots F_{i-1}^2 F_{i+1}^2 \cdots F_N^2 \leq F_{j+1,j}^2 \cdots F_{i-1,j}^2 F_{i+1,i,j}^2 \cdots F_{N,i,j}^2 \quad (3.15)$$

and

$$\begin{aligned} F_j^2 \cdots F_N^2 &\geq F_{j+1,j}^2 \cdots F_{i-1,j}^2 F_{i+1,i,j}^2 \cdots F_{N,i,j}^2 \\ &\times \left(1 - \sum_{k=1, k \neq i,j}^N [1 - f(|\mathbf{x}_j - \mathbf{x}_k|)^2]\right) \\ &\times \left(1 - \sum_{k=1, k \neq i}^N [1 - f(|\mathbf{x}_i - \mathbf{x}_k|)^2]\right). \end{aligned} \quad (3.16)$$

We now consider the first two terms in Eq. (3.11). In the numerator of the first term for each fixed i we use the estimate

$$f'(t_i)^2 \leq \sum_{j=1}^{i-1} f'(|\mathbf{x}_i - \mathbf{x}_j|)^2, \quad (3.17)$$

and in the second term we use $F_i \leq f(|\mathbf{x}_i - \mathbf{x}_j|)$. For fixed i and j one eliminates $\mathbf{x}_i$ and $\mathbf{x}_j$ from the rest of the integrand by using Eq. (3.15) and $F_j \leq 1$ in the numerator and Eq. (3.16) in the denominator to do the $\mathbf{x}_i$ and $\mathbf{x}_j$ integrations. With the transformation $\boldsymbol{\eta} = \mathbf{x}_i - \mathbf{x}_j$, $\boldsymbol{\chi} = (\mathbf{x}_i + \mathbf{x}_j)/2$ one gets

$$\begin{aligned} &\int [2f'(|\mathbf{x}_i - \mathbf{x}_j|)^2 + v(|\mathbf{x}_i - \mathbf{x}_j|)f(|\mathbf{x}_i - \mathbf{x}_j|)^2] \\ &\times g(\mathbf{x}_i)^2 g(\mathbf{x}_j)^2 d\mathbf{x}_i d\mathbf{x}_j \\ &= \int [2f'(|\boldsymbol{\eta}|)^2 + v(|\boldsymbol{\eta}|)f(|\boldsymbol{\eta}|)^2] \\ &\times g\left(\boldsymbol{\chi} + \frac{1}{2}\boldsymbol{\eta}\right)^2 g\left(\boldsymbol{\chi} - \frac{1}{2}\boldsymbol{\eta}\right)^2 d\boldsymbol{\eta} d\boldsymbol{\chi}. \end{aligned} \quad (3.18)$$

By the Cauchy-Schwarz inequality

$$\int g\left(\boldsymbol{\chi} + \frac{1}{2}\boldsymbol{\eta}\right)^2 g\left(\boldsymbol{\chi} - \frac{1}{2}\boldsymbol{\eta}\right)^2 d\boldsymbol{\chi} \leq \int g(\boldsymbol{\chi})^4 d\boldsymbol{\chi}, \quad (3.19)$$

and one obtains

$$\begin{aligned} &\int [2f'(|\mathbf{x}_i - \mathbf{x}_j|)^2 + v(|\mathbf{x}_i - \mathbf{x}_j|)f(|\mathbf{x}_i - \mathbf{x}_j|)^2] \\ &\times g(\mathbf{x}_i)^2 g(\mathbf{x}_j)^2 d\mathbf{x}_i d\mathbf{x}_j \\ &\leq \int g(\boldsymbol{\chi})^4 d\boldsymbol{\chi} \int [2f'(|\boldsymbol{\eta}|)^2 + v(|\boldsymbol{\eta}|)f(|\boldsymbol{\eta}|)^2] d\boldsymbol{\eta} \\ &= 2 \int g(\boldsymbol{\chi})^4 d\boldsymbol{\chi} J. \end{aligned} \quad (3.20)$$

In the denominator one gets, using that $0 \leq g \leq 1$,

$$\begin{aligned} &\int \left(1 - \sum_{p=1, p \neq i}^N [1 - f(|\mathbf{x}_p - \mathbf{x}_i|)^2]\right) g(\mathbf{x}_i)^2 d\mathbf{x}_i \\ &\geq \int g(\mathbf{x})^2 d\mathbf{x} - N \int [1 - f(|\mathbf{x}_p - \mathbf{x}_i|)^2] \\ &= \int g(\mathbf{x})^2 d\mathbf{x} - NI. \end{aligned} \quad (3.21)$$

The same factor comes from the $\mathbf{x}_j$ integration, the remaining factors are identical in numerator and denominator, and so finally the first and second term are bounded by

$$\sum_{i=1}^N (i-1) \frac{2 \int g(\mathbf{x})^4 d\mathbf{x} J}{\left[\int g(\mathbf{x})^2 d\mathbf{x} - NI\right]^2} \leq N^2 \frac{\int g(\mathbf{x})^4 d\mathbf{x} J}{\left[\int g(\mathbf{x})^2 d\mathbf{x} - NI\right]^2}. \quad (3.22)$$

A similar argument is now applied to the third term of Eq. (3.11). Note that the contributions from $k=i$ and $k < i$ are the same. Therefore

$$\begin{aligned} &\sum_{k=1}^i \int |\epsilon_{ik}| f(t_i) f'(t_i) |\epsilon_{jk}| f(t_j) f'(t_j) g(\mathbf{x}_i)^2 g(\mathbf{x}_j)^2 d\mathbf{x}_i d\mathbf{x}_j \\ &\leq 2 \sum_{k=i}^{i-1} \int f(|\mathbf{x}_i - \mathbf{x}_k|) f'(|\mathbf{x}_i - \mathbf{x}_k|) f(|\mathbf{x}_j - \mathbf{x}_k|) \\ &\times f'(|\mathbf{x}_j - \mathbf{x}_k|) g(\mathbf{x}_i)^2 g(\mathbf{x}_j)^2 d\mathbf{x}_i d\mathbf{x}_j. \end{aligned} \quad (3.23)$$

With $g \leq 1$ one gets

$$\begin{aligned} &2 \sum_{k=1}^{i-1} \int f(|\mathbf{x}_i|) f'(|\mathbf{x}_i|) f(|\mathbf{x}_j|) f'(|\mathbf{x}_j|) d\mathbf{x}_i d\mathbf{x}_j \\ &= 2(i-1) \left(\int f(|\mathbf{x}|) f'(|\mathbf{x}|) d\mathbf{x} \right)^2 \\ &= 2(i-1) K^2. \end{aligned} \quad (3.24)$$

The summation over i and j gives

$$\sum_{j=1}^N \sum_{i=1}^{j-1} (i-1) = \frac{1}{6} N(N-1)(N-2). \quad (3.25)$$

In the denominator one gets the same factors as above, and so the third term is bounded by

$$\frac{2}{3} N^3 \frac{K^2}{\left[\int g(\mathbf{x})^2 d\mathbf{x} - NI \right]^2}. \quad (3.26)$$

Next consider the last term of Eq. (3.11). Define $\tilde{e}$ by

$$\int [-g(\mathbf{x}) \nabla^2 g(\mathbf{x}) + V(\mathbf{x}) g(\mathbf{x})^2] d\mathbf{x} \equiv \tilde{e} \int g(\mathbf{x})^2 d\mathbf{x}. \quad (3.27)$$

After eliminating $\mathbf{x}_i$ from the integrands in the numerator and the denominator and using $F_i \leq 1$ one sees that the term is bounded above by

$$\frac{\tilde{e} \int g(\mathbf{x})^2 d\mathbf{x}}{\int g(\mathbf{x})^2 d\mathbf{x} - NI}. \quad (3.28)$$

Putting all terms together we obtain as an upper bound for the ground-state energy:

$$\begin{aligned} \frac{E^{\text{QM}}}{N} &\leq \frac{\tilde{e} \int g(\mathbf{x})^2 d\mathbf{x}}{\int g(\mathbf{x})^2 d\mathbf{x} - NI} + N \frac{\int g(\mathbf{x})^4 d\mathbf{x} J}{\left[\int g(\mathbf{x})^2 d\mathbf{x} - NI \right]^2} \\ &\quad + \frac{2}{3} N^2 \frac{K^2}{\left[\int g(\mathbf{x})^2 d\mathbf{x} - NI \right]^2} \end{aligned} \quad (3.29)$$

with I , J , K , and $\tilde{e}$ defined by Eqs. (3.21), (3.20), (3.24), and (3.27). It remains to choose f . We take for $b > a$ (we shall soon fix b)

$$f(r) = \begin{cases} (1 + \epsilon)u(r)/r & \text{for } r \leq b \\ 1 & \text{for } r > b \end{cases} \quad (3.30)$$

where $u(r)$ is the solution of the scattering equation

$$-u''(r) + \frac{1}{2}v(r)u(r) = 0 \quad (3.31)$$

with

$$u(0) = 0, \quad \lim_{r \rightarrow \infty} u'(r) = 1, \quad (3.32)$$

and ϵ is determined by requiring f to be continuous. Convexity of u gives

$$r \geq u(r) \geq \begin{cases} 0 & \text{for } r \leq a \\ r - a & \text{for } r > a, \end{cases}$$

$$1 \geq u'(r) \geq \begin{cases} 0 & \text{for } r \leq a \\ 1 - \frac{a}{r} & \text{for } r > a. \end{cases} \quad (3.33)$$

These estimates imply

$$J \leq (1 + \epsilon)^2 4\pi a, \quad (3.34)$$

$$I \leq 4\pi \left(\frac{a^3}{3} + ab(b-a) \right), \quad (3.35)$$

$$K \leq 4\pi (1 + \epsilon)a \left(b - \frac{a}{2} \right), \quad (3.36)$$

$$0 \leq \epsilon \leq \frac{a}{b-a}. \quad (3.37)$$

For Eq. (3.34) we used partial integration. By definition [Eq. (1.9)]

$$\bar{\rho} = \frac{1}{N} \int (\rho^{\text{GP}})^2 = N \frac{\int g^4}{\left(\int g^2 \right)^2}, \quad (3.38)$$

and we choose b such that

$$\frac{4\pi}{3} \bar{\rho} b^3 = \frac{\int g^4}{\int g^2} = \frac{\bar{\rho}}{\|\rho^{\text{GP}}\|_\infty} \equiv c. \quad (3.39)$$

With this choice the factor in the denominators in Eq. (3.29) is bounded by

$$\int g^2 - NI \geq \int g^2 \left(1 - \frac{a}{b} \right)^3. \quad (3.40)$$

Note that $c \leq 1$, and $a < b$ holds provided

$$\frac{a^3}{b^3} = \frac{4\pi}{3} a^3 \|\rho^{\text{GP}}\|_\infty < 1. \quad (3.41)$$

Collecting the estimates Eqs. (3.34)–(3.37), we finally obtain the following theorem.

Theorem III.1 (Upper bound for the QM energy).

$$E^{\text{QM}} \leq \frac{\int [|\nabla \Phi^{\text{GP}}|^2 + V(\Phi^{\text{GP}})^2]}{\left(1 - \frac{a}{b}\right)^3} + 4\pi a \int (\Phi^{\text{GP}})^4 \frac{1 + \frac{2}{c} \left(\frac{a}{b}\right) - \frac{2}{c} \left(\frac{a}{b}\right)^2 + \frac{1}{2c} \left(\frac{a}{b}\right)^3}{\left(1 - \frac{a}{b}\right)^8} \quad (3.42)$$

with b and c defined by Eq. (3.39).

Remark III.1 (Negative potentials with hard core). Equation (3.1) can be extended to include partially negative potentials of the form

$$v(r) = \begin{cases} \infty & \text{for } 0 \leq r \leq d \\ -|w(r)| & \text{for } d < r \leq R_0 \\ 0 & \text{for } r > R_0, \end{cases} \quad (3.43)$$

as long as $f'(r)^2 + \frac{1}{2}v(r)f(r)^2 \geq 0$ for all r . With f from Eq. (3.30), this is the case for sufficiently shallow potentials. The potential energy is then negative, and the estimates used for Eq. (3.22) are no longer valid. But

$$\sum_{j < i} v(|\mathbf{x}_i - \mathbf{x}_j|) F_i^2 \leq \sum_i v(t_i) F_i^2, \quad (3.44)$$

and because of $2f'(r)^2 + v(r)f(r)^2 \geq 0$ we get

$$2f'(t_i)^2 + v(t_i)f(t_i)^2 \leq \sum_{j=1}^{i-1} 2f'(|\mathbf{x}_i - \mathbf{x}_j|)^2 + v(|\mathbf{x}_i - \mathbf{x}_j|)f(|\mathbf{x}_i - \mathbf{x}_j|)^2, \quad (3.45)$$

which replaces Eq. (3.17). Note that for potentials as in Eq. (3.43) f satisfies Eq. (3.5), as long as $a > 0$.

Remark III.2 (Homogeneous gas). For the special case of a homogeneous Bose gas (i.e., $V=0$) in a box of volume $\mathcal{V}$, the GP density is simply

$$\rho^{\text{GP}}(\mathbf{x}) = N/\mathcal{V} = \bar{\rho}, \quad (3.46)$$

for all $\mathbf{x}$ in the box, and the GP energy is given by

$$E^{\text{GP}}(N, a) = 4\pi a \frac{N^2}{\mathcal{V}}. \quad (3.47)$$

Our method also applies here, if we impose periodic boundary conditions on the box. Therefore our upper bound is a generalization of a result by Dyson [7], who proved an analogous bound for the special case of a homogeneous Bose gas with hard-sphere interaction.

B. Upper bound for the chemical potential

By the same method as in the previous subsection one can derive a bound on the increase of the energy when one particle is added to the system. This bound will be needed for the derivation of the lower bound to the energy.

Theorem III.2 (Upper bound for the chemical potential). Let $E^*(N, a)$ denote the infimum of the functional

$$\mathcal{E}^*[\Phi] = \int [|\nabla \Phi(\mathbf{x})|^2 + V(\mathbf{x})|\Phi(\mathbf{x})|^2 + 8\pi a \|\Phi\|_\infty^2 |\Phi(\mathbf{x})|^2] d\mathbf{x} \quad (3.48)$$

with $\int |\Phi|^2 = N$. Let Φ^* be the positive minimizer of $\mathcal{E}^*$ (its existence is guaranteed by the same arguments as for the GP functional itself), and $\bar{\rho}^* = \int \Phi^{*4}/N$. Then

$$E^{\text{QM}}(N+1, a) \leq E^{\text{QM}}(N, a) + E^*(1, Na) [1 + O(a\bar{\rho}^{*1/3})]. \quad (3.49)$$

Proof. Let $\Psi_0^{(N)}$ be the ground-state wave function of $H^{(N)}$. As test wave function for $H^{(N+1)}$ we take

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_{N+1}) = \Psi_0^{(N)}(\mathbf{x}_1, \dots, \mathbf{x}_N) \Phi^*(\mathbf{x}_{N+1}) f(t_{N+1}), \quad (3.50)$$

where f and t_{N+1} are defined as in Eqs. (3.30) and (3.4), i.e., $rf(r)$ is essentially the zero-energy scattering solution and t_{N+1} is the distance of $\mathbf{x}_{N+1}$ from its nearest neighbor. We have

$$\begin{aligned} \langle \Psi | H^{(N+1)} | \Psi \rangle &= \int f^2 \Phi^{*2} \left(\sum_{i=1}^N (-\Psi_0^{(N)} \nabla_i^2 \Psi_0^{(N)} + V(\mathbf{x}_i) |\Psi_0^{(N)}|^2) + \sum_{i < j}^N v(|\mathbf{x}_i - \mathbf{x}_j|) |\Psi_0^{(N)}|^2 \right) + \int |\Psi_0^{(N)}|^2 \Phi^{*2} \left(\sum_{i=1}^N |\nabla_i f|^2 \right) \\ &\quad + \int |\Psi_0^{(N)}|^2 f^2 [-\Phi^* \nabla_{N+1}^2 \Phi^* + V(\mathbf{x}_{N+1}) \Phi^{*2}] + \int |\Psi_0^{(N)}|^2 \Phi^{*2} \left(|\nabla_{N+1} f|^2 + \sum_{i=1}^N v(|\mathbf{x}_{N+1} - \mathbf{x}_i|) f^2 \right). \end{aligned} \quad (3.51)$$

For f one uses the estimates

$$f \leq 1, \quad f(t_{N+1})^2 \geq 1 - \sum_{i=1}^N [1 - f(|\mathbf{x}_{N+1} - \mathbf{x}_i|)^2], \quad (3.52)$$

and for the derivatives one has

$$|\nabla_{N+1}f|^2 = f'(t_{N+1})^2 = \sum_{i=1}^N |\nabla_i f|^2 \quad (3.53)$$

and

$$f'(t_{N+1})^2 \leq \sum_{i=1}^N f'(|\mathbf{x}_{N+1} - \mathbf{x}_i|)^2. \quad (3.54)$$

After division by the norm of Ψ Eq. (3.51) becomes

$$\begin{aligned} E^{\text{QM}}(N+1, a) \leq E^{\text{QM}}(N, a) &+ \frac{\int |\Psi_0^{(N)}|^2 (-\Phi^* \nabla_{N+1}^2 \Phi^* + V \Phi^* \Phi^2)}{\int |\Psi_0^{(N)}|^2 \left[N - N \Phi^* \Phi^2 \int (1 - f^2) \right]} \\ &+ \frac{\int |\Psi_0^{(N)}|^2 \|\Phi^*\|_\infty^2 \sum_{i=1}^N [2f'(|\mathbf{x}_{N+1} - \mathbf{x}_i|)^2 + v(|\mathbf{x}_{N+1} - \mathbf{x}_i|)f(|\mathbf{x}_{N+1} - \mathbf{x}_i|)^2]}{\int |\Psi_0^{(N)}|^2 \left[N - N \|\Phi^*\|_\infty^2 \int (1 - f^2) \right]}. \end{aligned} \quad (3.55)$$

$\Psi_0^{(N)}$ does not depend on $\mathbf{x}_{N+1}$. One integrates first over $\mathbf{x}_{N+1}$ and then over the remaining variables. In analogy with the estimates (3.34) and (3.40) one gets

$$\int [2f'(|\mathbf{x}|)^2 + v(|\mathbf{x}|)f(|\mathbf{x}|)^2] d\mathbf{x} \leq 8\pi a [1 + O(a\bar{\rho}^{*1/3})] \quad (3.56)$$

and

$$\|\Phi^*\|_\infty^2 \int [1 - f(|\mathbf{x}|)^2] d\mathbf{x} \leq O(a\bar{\rho}^{*1/3}). \quad (3.57)$$

This implies

$$E^{\text{QM}}(N+1, a) \leq E^{\text{QM}}(N, a) + \frac{E^*(N, a)}{N} [1 + O(a\bar{\rho}^{*1/3})]. \quad (3.58)$$

By scaling, $E^*(N, a) = NE^*(1, Na)$ and Eq. (3.49) follows. $\square$

Remark III.3 (Box with Neumann conditions). Equation (3.49) also holds for a box with Neumann boundary conditions. To see this we note that Neumann conditions give the lowest energy for the quadratic form $\langle \Psi | H^{(N+1)} | \Psi \rangle$; therefore it is possible to use Eq. (3.50) as a test function, even if f does not fulfill Neumann conditions. If $\Psi_0^{(N)}$ and Φ^* do, the calculation above is still valid, since for

$$\int |\nabla(gf)|^2 = - \int f^2 g \nabla^2 g + \int g^2 |\nabla f|^2 \quad (3.59)$$

only boundary conditions for g are needed.

Note also that in the homogeneous case, i.e., $V=0$ in the box, $E^*(N, a) = 2E^{\text{GP}}(N, a)$.

IV. LOWER BOUNDS

A. The homogeneous case

In [5] the following lower bound was established for the ground-state energy, E^{hom} , of a Bose gas in a box of side length L with Neumann boundary conditions and v of finite range:

$$E^{\text{hom}}(N, L) \geq 4\pi a \frac{N^2}{L^3} (1 - CY^{1/17}) \quad (4.1)$$

with $Y = a^3 N/L^3$ and C a constant. The estimate holds for all Y small enough and $L/a \gg Y^{-6/17}$ (note that this implies $N \gg Y^{-1/17}$). In the thermodynamic limit the constant C can be taken to be $C=8.9$, but this value is only of academic interest, because the error term $-CY^{1/17}$ is not believed to reflect the true state of affairs. Presumably, it does not even have the correct sign.

The restriction of a finite range can be relaxed. In fact, Eq. (4.1) holds (with a different constant C) for all positive, spherically symmetric v with

$$v(r) \leq \text{const} \times r^{-(3+\epsilon+1/5)} \quad \text{for } r \text{ large enough, } \epsilon > 0. \quad (4.2)$$

More generally, if

$$v(r) \leq \text{const} \times r^{-(3+\epsilon)} \quad \text{for } r \text{ large enough, } \epsilon > 0, \quad (4.3)$$

then Eq. (4.1) holds at least with the exponent $1/17$ replaced by an exponent $O(\epsilon)$. We prove these assertions in Appendix B.

In the next section we shall stick to the estimate (4.1) for simplicity, but in the limit $N \rightarrow \infty$ the explicit form of the error term is not significant so a decrease of the potential as in Eq. (4.3) is sufficient for the limit Theorems I.1 and I.2.

B. The lower bound in the inhomogeneous case

Our generalization of Eq. (4.1) to the inhomogeneous case is as follows:

Theorem IV.1 (Lower bound for the QM energy). Let v be positive, spherically symmetric, and decrease at infinity like Eq. (4.2). Its scattering length is $a = a_1/N$ with a_1 fixed, as explained in the Introduction. Then as $N \rightarrow \infty$

$$E_R^{\text{QM}}(N, a) \geq E_R^{\text{GP}}(N, a) (1 - \text{const} \times N^{-1/10}) \quad (4.4)$$

for all R large enough, where E_R^{GP} is the GP energy in a cube with side length $2R$, center at the origin, and Neumann boundary conditions; the constant in Eq. (4.4) depends only on a_1 and R .

Proof. As in [5] the lower bound will be obtained by dividing space into cubic boxes with Neumann conditions at the boundary, which only lowers the energy. Moreover, interactions among particles in different boxes are dropped. Since $v \geq 0$, this, too, lowers the energy. For the lower bound one has to estimate the energy for a definite particle number in each box and then optimize over all distributions of the N particles among the boxes.

Step 1 (Finite box): The first step is to show that all the particles can be assumed to be in some large but finite box. Since

$$K(R) = \inf_{|x| > R} V(x) \quad (4.5)$$

tends monotonically to ∞ with R , one knows that the energy of a particle outside a cube Λ_R of side length $2R$ and center at the origin is at least $K(R)$. Hence

$$E^{\text{QM}}(N, a) \geq \inf_{0 \leq n \leq N} \{E_R^{\text{QM}}(N-n, a) + nK(R)\}, \quad (4.6)$$

where $E_R^{\text{QM}}(N-n, a)$ denotes the energy of $N-n$ particles in Λ_R , with Neumann conditions at the boundary. We now apply Theorem III.2 (which holds also in a cube with Neumann conditions). Applying the theorem n times and noting that $E^*(1, Na)$ is monotone in N we have

$$E_R^{\text{QM}}(N-n, a) \geq E_R^{\text{QM}}(N, a) - nE^*(1, Na)[1 + O(a\bar{\rho}^{*1/3})]. \quad (4.7)$$

Hence there is a constant K' (that depends only on Na), such that $K(R) > K'$ implies that the infimum is obtained at $n=0$. This is fulfilled for all sufficiently large R , independently of N if Na is fixed. So we can restrict ourselves to estimating the energy in Λ_R with Neumann boundary conditions.

Step 2 (Trading V for $-\rho_R^{\text{GP}}$): We shall now use the GP equation to eliminate V from the problem, effectively replacing it by $-8\pi a\rho_R^{\text{GP}}$. We write the wave function in Λ_R^N as

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = f(\mathbf{x}_1, \dots, \mathbf{x}_N) \prod_{i=1}^N \Phi_R^{\text{GP}}(\mathbf{x}_i), \quad (4.8)$$

where Φ_R^{GP} denotes the the minimizer of the GP equation in Λ_R ; since it is strictly positive, every wave function can be written in this form. Note also that Φ_R^{GP} and f obey Neumann conditions. We have

$$\begin{aligned} \langle \Psi | H \Psi \rangle &= \sum_{i=1}^N \int |\Psi|^2 \Phi_R^{\text{GP}}(\mathbf{x}_i)^{-1} \\ &\quad \times [-\nabla_i^2 + V(\mathbf{x}_i)] \Phi_R^{\text{GP}}(\mathbf{x}_i) d^N \mathbf{x} \\ &\quad + \sum_{i=1}^N \int \prod_{k=1}^N \Phi_R^{\text{GP}}(\mathbf{x}_k)^2 |\nabla_i f|^2 d^N \mathbf{x} \\ &\quad + \sum_{i < j}^N \int |\Psi|^2 v(|\mathbf{x}_i - \mathbf{x}_j|) d^N \mathbf{x}, \end{aligned} \quad (4.9)$$

where the integrals are over Λ_R^N . Using the GP equation (2.4) this becomes

$$\begin{aligned} \langle \Psi | H \Psi \rangle &= \sum_{i=1}^N \int \prod_{k=1}^N \rho_R^{\text{GP}}(\mathbf{x}_k) \left[|\nabla_i f|^2 \right. \\ &\quad \left. + \left(\mu_R - 8\pi a \rho_R^{\text{GP}}(\mathbf{x}_i) + \sum_{j=1}^{i-1} v(|\mathbf{x}_i - \mathbf{x}_j|) \right) |f|^2 \right]. \end{aligned} \quad (4.10)$$

Inserting the value (2.5) for μ_R gives

$$\frac{\langle \Psi | H \Psi \rangle}{\langle \Psi | \Psi \rangle} - E_R^{\text{GP}} = 4\pi a \bar{\rho}_R N + Q(f) \quad (4.11)$$

with

$$Q(f) = \frac{\sum_{i=1}^N \int_{\Lambda_R^N} \prod_{k=1}^N \rho_R^{\text{GP}}(\mathbf{x}_k) \left(|\nabla_i f|^2 + \sum_{j=1}^{i-1} v(|\mathbf{x}_i - \mathbf{x}_j|) |f|^2 - 8\pi a \rho_R^{\text{GP}}(\mathbf{x}_i) |f|^2 \right)}{\int_{\Lambda_R^N} \prod_{k=1}^N \rho_R^{\text{GP}}(\mathbf{x}_k) |f|^2}. \quad (4.12)$$

Step 3 (division into boxes): $Q(f)$ is a normalized quadratic form on the weighted L^2 space $L^2(\Lambda_R^N, \prod_{k=1}^N \rho_R^{\text{GP}}(\mathbf{x}_k) d\mathbf{x}_k)$, and

can be minimized by dividing the cube Λ_R into smaller cubes with side length L , labeled by α , distributing the N particles among the boxes, and optimizing over all distributions. We therefore have

$$\inf_f Q(f) \geq \inf_{\{n_\alpha\}} \sum_\alpha \inf_{f_\alpha} Q_\alpha(f_\alpha), \quad (4.13)$$

where the infimum is taken over all distributions of the particles with $\sum n_\alpha = N$, and $Q_\alpha(f)$ is given by

$$Q_\alpha(f) = \sum_{i=1}^{n_\alpha} \frac{\int \prod_{k=1}^{n_\alpha} \rho_R^{\text{GP}}(\mathbf{x}_k) \left(|\nabla_i f|^2 + \sum_{j=1}^{i-1} v(|\mathbf{x}_i - \mathbf{x}_j|) |f|^2 - 8\pi a \rho_R^{\text{GP}}(\mathbf{x}_i) |f|^2 \right)}{\int \prod_{k=1}^{n_\alpha} \rho_R^{\text{GP}}(\mathbf{x}_k) |f|^2}, \quad (4.14)$$

where the integrals are over $\mathbf{x}_k$ in the box α , $k = 1, \dots, n_\alpha$. Note that here $f = f(\mathbf{x}_1, \dots, \mathbf{x}_{n_\alpha})$, and Eq. (4.14) is the same as Eq. (4.12) with N replaced by n_α and Λ_R with the box α .

We now want to use Eq. (4.1) and therefore we must approximate ρ_R^{GP} by constants in each box. Let $\rho_{\alpha, \max}$ and $\rho_{\alpha, \min}$, respectively, denote the maximal and minimal values of ρ_R^{GP} in box α . With

$$\Phi^{(i)}(\mathbf{x}_1, \dots, \mathbf{x}_{n_\alpha}) = f(\mathbf{x}_1, \dots, \mathbf{x}_{n_\alpha}) \prod_{k=1, k \neq i}^{n_\alpha} \Phi_R^{\text{GP}}(\mathbf{x}_k), \quad (4.15)$$

one has

$$\begin{aligned} & \frac{\int \prod_k \rho_R^{\text{GP}}(\mathbf{x}_k) \left(|\nabla_i f|^2 + \sum_{j=1}^{i-1} v(|\mathbf{x}_i - \mathbf{x}_j|) |f|^2 \right)}{\int \prod_k \rho_R^{\text{GP}}(\mathbf{x}_k) |f|^2} \\ & \geq \frac{\rho_{\alpha, \min}}{\rho_{\alpha, \max}} \frac{\int |\nabla_i \Phi^{(i)}|^2 + \sum_{j=1}^{i-1} v(|\mathbf{x}_i - \mathbf{x}_j|) |\Phi^{(i)}|^2}{\int |\Phi^{(i)}|^2}. \end{aligned} \quad (4.16)$$

This holds for all i , and if we use $\rho_R^{\text{GP}}(\mathbf{x}_i) \leq \rho_{\alpha, \max}$ in Eq. (4.14), we get

$$Q_\alpha(f) \geq \frac{\rho_{\alpha, \min}}{\rho_{\alpha, \max}} E^{\text{hom}}(n_\alpha, L) - 8\pi a \rho_{\alpha, \max} n_\alpha, \quad (4.17)$$

where E^{hom} is the energy in a box without an external potential.

Remark: If we had not taken Step 2 and used instead the division into boxes directly on the original Hamiltonian (1.1) we would be considering the minimization of

$$\sum_\alpha E^{\text{hom}}(n_\alpha, L) + V_{\alpha, \min} n_\alpha. \quad (4.18)$$

Such a procedure, however, would not lead to the GP energy. To see this, consider the special case of no interaction, i.e., $v = 0$ and hence also $E^{\text{hom}}(n_\alpha, L) = 0$. The minimum of Eq. (4.18) is then simply $N \min_{\mathbf{x}} V(\mathbf{x})$, whereas the GP energy is in this case N times the ground-state energy of $-\nabla^2 + V$.

Step 4 (Minimizing in each box): Dropping the subsidiary condition $\sum n_\alpha = N$ can only lower the infimum. Hence it is sufficient to minimize each Q_α separately. To justify the use of Eq. (4.1), we have to ensure that n_α is large enough. But if the minimum is taken for some $\bar{n}_\alpha$, we have

$$\frac{\rho_{\alpha, \min}}{\rho_{\alpha, \max}} [E^{\text{hom}}(\bar{n}_\alpha + 1, L) - E^{\text{hom}}(\bar{n}_\alpha, L)] \geq 8\pi a \rho_{\alpha, \max}, \quad (4.19)$$

and using Theorem III.2, which states that

$$E^{\text{hom}}(\bar{n}_\alpha + 1, L) - E^{\text{hom}}(\bar{n}_\alpha, L) \leq 8\pi a \frac{\bar{n}_\alpha}{L^3} [1 + O(\bar{n}_\alpha a^3/L^3)], \quad (4.20)$$

we see that $\bar{n}_\alpha$ is at least $\sim \rho_{\alpha, \max} L^3$. We shall later choose $L \sim N^{-1/10}$, so the conditions needed for Eq. (4.1) are fulfilled for N large enough, since $\rho_{\alpha, \max} \sim N$ and hence $\bar{n}_\alpha \sim N^{7/10}$, $L/a \sim N^{9/10}$, and $Y_\alpha \sim N^{-2}$. Thus we have (for large enough N)

$$Q_\alpha(f) \geq 4\pi a \left(\frac{\rho_{\alpha, \min}}{\rho_{\alpha, \max}} \frac{n_\alpha^2}{L^3} (1 - CY_\alpha^{1/17}) - 2n_\alpha \rho_{\alpha, \max} \right). \quad (4.21)$$

We now use $Y_\alpha = a^3 n_\alpha / L^3 \leq a^3 N / L^3 = Y$, and drop the requirement that n_α has to be an integer. The minimum of Eq. (4.21) is obtained for

$$n_\alpha = \frac{\rho_{\alpha, \max}^2}{\rho_{\alpha, \min}} \frac{L^3}{(1 - CY^{1/17})}. \quad (4.22)$$

This gives for Eq. (4.11)

$$E^{\text{QM}}(N, a) - E_R^{\text{GP}}(N, a)$$

$$\geq 4\pi a \bar{\rho}_R N - 4\pi a \sum_{\alpha} \rho_{\alpha, \min}^2 L^3 \left(\frac{\rho_{\alpha, \max}^3}{\rho_{\alpha, \min}^3 (1 - CY^{1/17})} \right). \quad (4.23)$$

Now ρ_R^{GP} is differentiable by Lemma A.6, and strictly positive. Since all the boxes are in the fixed cube Λ_R there are constants $C' < \infty$, $C'' > 0$, such that

$$\rho_{\alpha, \max} - \rho_{\alpha, \min} \leq NC' L, \quad \rho_{\alpha, \min} \geq NC''. \quad (4.24)$$

Therefore we have, for Y and L small,

$$\frac{\rho_{\alpha, \max}^3}{\rho_{\alpha, \min}^3 (1 - CY^{1/17})} \leq 1 + DY^{1/17} + D' L \quad (4.25)$$

with suitable constants D and D' . Also,

$$4\pi a \sum_{\alpha} \rho_{\alpha, \min}^2 L^3 \leq 4\pi a \int (\rho_R^{\text{GP}})^2 \leq E_R^{\text{GP}}(N, a), \quad (4.26)$$

and hence

$$E^{\text{QM}}(N, a) \geq E_R^{\text{GP}}(N, a) (1 - DY^{1/17} - D' L). \quad (4.27)$$

As a last step it remains to optimize the length L . Recall that $Y = a^3 N / L^3$ and Na is fixed. The exponents of N in both error terms in Eq. (4.27) are the same for

$$L \sim a N^{9/10} \sim N^{-1/10}. \quad (4.28)$$

The final result, therefore, is

$$E^{\text{QM}}(N, a) \geq E_R^{\text{GP}}(N, a) (1 - D'' N^{-1/10}). \quad (4.29)$$

□

C. The limit theorems

By Theorems III.1 and IV.1 we have (with $a = a_1 / N$)

$$\begin{aligned} E^{\text{GP}}(N, a) [1 + O(N^{-2/3})] &\geq E^{\text{QM}}(N, a) \\ &\geq E_R^{\text{GP}}(N, a) [1 - O(N^{-1/10})]. \end{aligned} \quad (4.30)$$

Dividing by N and taking the limit $N \rightarrow \infty$ we have

$$E^{\text{GP}}(1, a_1) \geq \lim_{N \rightarrow \infty} \frac{E^{\text{QM}}(N, a_1 / N)}{N} \geq E_R^{\text{GP}}(1, a_1) \quad (4.31)$$

for all R large enough. Using Lemma II.2 and taking the limit $R \rightarrow \infty$ we finally prove Theorem I.1.

The convergence of the energy implies also the convergence of the densities: We replace V by $V + \delta W$ with $W \in L^\infty$ and $\delta \in \mathbb{R}$, and denote the corresponding energies by $E_\delta(N, a)$. It is no restriction to assume that $V + \delta W \geq 0$ for small $|\delta|$. $E_\delta^{\text{QM}}(N, a_1 / N) / N$ is concave in δ (it is an infimum over linear functions), and converges for each δ to $E_\delta^{\text{GP}}(1, a_1)$

as $N \rightarrow \infty$. This implies convergence of the derivatives and we have (Feynman-Hellmann principle)

$$\begin{aligned} \frac{\partial}{\partial \delta} E_\delta^{\text{QM}}(N, a) \Big|_{\delta=0} &= \int_{\mathbb{R}^3} W \rho_{N, a}^{\text{QM}}, \\ \frac{\partial}{\partial \delta} E_\delta^{\text{GP}}(N, a) \Big|_{\delta=0} &= \int_{\mathbb{R}^3} W \rho_{N, a}^{\text{GP}} \end{aligned} \quad (4.32)$$

with $\rho_{N, a}^{\text{QM}}$ given by Eq. (1.3). In the weak L^1 sense we thus have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \rho_{N, a_1 / N}^{\text{QM}}(\mathbf{x}) = \rho_{1, a_1}^{\text{GP}}(\mathbf{x}), \quad (4.33)$$

which proves Theorem 1.2.

V. CONCLUSIONS

We have proved that the GP energy functional correctly describes the energy and particle density of a Bose gas in a trap to leading order in the small parameter ρa^3 (where $\bar{\rho}$ is the mean density and a is the scattering length) in the limit where the particle number N tends to infinity, but a tends to zero with Na fixed.

ACKNOWLEDGMENT

The work was partially supported by the National Science Foundation, Grant No. PHY 98-20650.

APPENDIX A

In this appendix we prove Theorem II.1. The proof is split into several lemmas.

Lemma A.1 (Strict convexity). For $\rho \geq 0$, $\sqrt{\rho} \in \mathcal{D}$, $\mathcal{E}^{\text{GP}}[\sqrt{\rho}]$ is strictly convex in ρ .

Proof. The second term in Eq. (1.4) is linear, the third quadratic in ρ . So it suffices to show that the first term is convex. Let ρ_1 and ρ_2 be given, with $\Phi_1 = \rho_1^{1/2}$ and $\Phi_2 = \rho_2^{1/2}$ in $\mathcal{D}_N$. Then also $\Phi = [\alpha \rho_1 + (1 - \alpha) \rho_2]^{1/2} \in \mathcal{D}_N$ for all $0 < \alpha < 1$. We have

$$\begin{aligned} \Phi \nabla \Phi &= \alpha \Phi_1 \nabla \Phi_1 + (1 - \alpha) \Phi_2 \nabla \Phi_2 \\ &= (\alpha^{1/2} \Phi_1) (\alpha^{1/2} \nabla \Phi_1) + [(1 - \alpha)^{1/2} \Phi_2] \\ &\quad \times [(1 - \alpha)^{1/2} \nabla \Phi_2] \\ &\leq [\alpha \Phi_1^2 + (1 - \alpha) \Phi_2^2]^{1/2} [\alpha |\nabla \Phi_1|^2 \\ &\quad + (1 - \alpha) |\nabla \Phi_2|^2]^{1/2} \\ &= \Phi (\alpha |\nabla \Phi_1|^2 + (1 - \alpha) |\nabla \Phi_2|^2)^{1/2}. \end{aligned}$$

Hence

$$|\nabla \Phi|^2 \leq \alpha |\nabla \Phi_1|^2 + (1 - \alpha) |\nabla \Phi_2|^2. \quad \square$$

Remark. Because $V \geq 0$, $\mathcal{E}^{\text{GP}}[\Phi]$ is also convex in $\Phi \in \mathcal{D}$. But since the domain $\mathcal{D}_N$ is not convex, it is necessary to consider $\rho \mapsto \mathcal{E}^{\text{GP}}[\sqrt{\rho}]$.

Lemma A.2 (Minimizer). *For all N there exists a minimizing $\Phi_\infty \in \mathcal{D}_N$, with $\mathcal{E}^{\text{GP}}[\Phi_\infty] = E^{\text{GP}}(N, a)$. Moreover, $|\Phi_\infty|^2$ is unique.*

Proof. Let Φ_n be a minimizing sequence in $\mathcal{D}_N$, i.e., $\lim_{n \rightarrow \infty} \mathcal{E}^{\text{GP}}[\Phi_n] = E^{\text{GP}}$. It is clear that there exists a constant C , such that $\|\nabla \Phi_n\|_2 < C$, $\|\Phi_n\|_4 < C$ and $\int |\Phi_n|^2 V < C$ for all n (recall that V is non-negative). Hence the sequence belongs to a weakly compact set in L^4 , as well as in the Sobolev space $H^1 = \{\Phi : \|\Phi\|_2^2 + \|\nabla \Phi\|_2^2 < \infty\}$, and in the space L^2_V , defined by the L^2 norm $\|\Phi\|_V = [\int |\Phi(\mathbf{x})|^2 V(\mathbf{x}) d\mathbf{x}]^{1/2}$. Thus, there exists a $\Phi_\infty \in \mathcal{D}$ and a weakly convergent subsequence, again denoted by Φ_n , such that

$$\Phi_n \rightharpoonup \Phi_\infty \text{ in } L^2 \cap L^4 \cap L^2_V$$

$$\nabla \Phi_n \rightharpoonup \nabla \Phi_\infty \text{ in } L^2.$$

Because the L^4 norm, the Sobolev norm, and the L^2_V norm are all weakly lower semicontinuous, we have

$$\liminf_{n \rightarrow \infty} \mathcal{E}^{\text{GP}}[\Phi_n] \geq \mathcal{E}^{\text{GP}}[\Phi_\infty],$$

and it remains only to show that $\Phi_\infty \in \mathcal{D}_N$. Since $|\Phi_n|^2$ converges to $|\Phi_\infty|^2$ in L^1_{loc} it is clear that $\|\Phi_\infty\|_2^2 \leq N$. Moreover,

$$\int_B |\Phi_n|^2 \rightarrow \int_B |\Phi_\infty|^2 \leq \|\Phi_\infty\|_2^2$$

for all bounded regions B . If $\|\Phi_\infty\|_2^2 = N - \epsilon$ with $\epsilon > 0$, then there exists a constant M_B for all B , such that

$$\int_{\mathbb{R}^3 \setminus B} |\Phi_n|^2 \geq \epsilon$$

for all $n > M_B$. Since $\lim_{|\mathbf{x}| \rightarrow \infty} V(\mathbf{x}) = \infty$, this would imply $\int V |\Phi_n|^2 \rightarrow \infty$, which is impossible because Φ_n is a minimizing sequence for the functional $\mathcal{E}^{\text{GP}}$. Hence $\|\Phi_\infty\|_2^2 = N$.

The uniqueness of $|\Phi_\infty|^2$ follows immediately from strict convexity, Lemma A.1. $\square$

Lemma A.3 (GP equation). *Every minimizing Φ_∞ satisfies the Gross-Pitaevskii equation (2.4). Conversely, every solution to Eq. (2.4), with μ given by Eq. (2.5), is a minimizer for $\mathcal{E}^{\text{GP}}$.*

Proof. Pick a function $f \in C_0^\infty$. The stationarity of $\mathcal{E}^{\text{GP}}$ at Φ_∞ implies

$$\frac{\partial}{\partial \epsilon} (\mathcal{E}^{\text{GP}}[\Phi_\infty + \epsilon f] + \mu \|\Phi_\infty + \epsilon f\|_2^2) \Big|_{\epsilon=0} = 0$$

with a Lagrange parameter μ to take account of the subsidiary condition $\|\Phi\|_2^2 = N$. With f real valued one obtains

$$-\nabla^2 \text{Re } \Phi_\infty + V \text{Re } \Phi_\infty + 8\pi a |\Phi_\infty|^2 \text{Re } \Phi_\infty = \mu \text{Re } \Phi_\infty$$

and an analogous equation for $\text{Im } \Phi_\infty$ with f purely imaginary. The value of μ is obtained by multiplying the GP equation with Φ_∞ and integrating. By the same argument, every

solution Φ to the GP equation satisfies $\mathcal{E}^{\text{GP}}[\Phi] = E^{\text{GP}}$ and is thus a minimizer. $\square$

Lemma A.4 (Uniqueness). *The minimizing Φ_∞ is unique up to a constant phase factor. This factor can be chosen so that Φ_∞ is strictly positive.*

Proof. Since $\mathcal{E}^{\text{GP}}[|\Phi|] \leq \mathcal{E}^{\text{GP}}[\Phi]$ (by an analogous computation as in the proof of Lemma A.1), we know that $|\Phi_\infty|$ is a minimizer and hence a solution to the GP equation. It is thus an eigenstate of the Hamiltonian $H = -\nabla^2 + W$ with $W = V + 8\pi a |\Phi_\infty|^2$ (recall that $|\Phi_\infty|^2$ is unique), and since it is non-negative, it must be a ground state. Since Φ_∞ solves the same equation it is also a ground state. Now $W \in L^2_{\text{loc}}$ and $\lim_{|\mathbf{x}| \rightarrow \infty} W(\mathbf{x}) = \infty$, so the ground state of H is unique up to a phase and without zeros (see [6], Sec. XIII.47). $\square$

The unique strictly positive minimizer is denoted by Φ^{GP} .

Lemma A.5 (Exponential falloff). *For all $t > 0$ there exists an M_t , such that $\Phi^{\text{GP}}(\mathbf{x}) \leq M_t e^{-t|\mathbf{x}|}$. In particular, $\Phi^{\text{GP}} \in L^\infty$.*

Proof. Put $W = V + 8\pi a (\Phi^{\text{GP}})^2$ and let $t > 0$. The GP equation implies

$$(-\nabla^2 + t^2)\Phi^{\text{GP}} = -(W - \mu - t^2)\Phi^{\text{GP}}.$$

Using the Yukawa potential $Y_t(\mathbf{x}) = (4\pi|\mathbf{x}|)^{-1} \exp(-t|\mathbf{x}|)$ we can rewrite this as

$$\Phi^{\text{GP}}(\mathbf{x}) = - \int Y_t(\mathbf{x} - \mathbf{y}) [W(\mathbf{y}) - \mu - t^2] \Phi^{\text{GP}}(\mathbf{y}) d\mathbf{y}.$$

Since $\Phi^{\text{GP}} > 0$, and $W(\mathbf{y}) - \mu - t^2 > 0$ for $|\mathbf{y}| > R$ with R large enough, we also have

$$\Phi^{\text{GP}}(\mathbf{x}) \leq - \int_{|\mathbf{y}| < R} Y_t(\mathbf{x} - \mathbf{y}) [W(\mathbf{y}) - \mu - t^2] \Phi^{\text{GP}}(\mathbf{y}) d\mathbf{y}.$$

Now $W\Phi^{\text{GP}} \in L^2_{\text{loc}}$, and hence

$$M_t = \sup_{\mathbf{x}} \int_{|\mathbf{y}| < R} \frac{\exp\{t(|\mathbf{x}| - |\mathbf{x} - \mathbf{y}|)\}}{4\pi|\mathbf{x} - \mathbf{y}|} \times [W(\mathbf{y}) - \mu - t^2] \Phi^{\text{GP}}(\mathbf{y}) d\mathbf{y} < \infty.$$

$\square$

Lemma A.6 (Regularity). *$\Phi^{\text{GP}}(\mathbf{x})$ is once continuously differentiable in $\mathbf{x} \in \mathbb{R}^3$, and $\nabla \Phi^{\text{GP}}$ is Hölder continuous of order 1. If $V \in C^\infty$, then $\Phi^{\text{GP}} \in C^\infty$. Moreover, $E^{\text{GP}}(N, a)$ is continuously differentiable in a and hence in N (by Eq. (1.6)), and $dE^{\text{GP}}(N, a)/dN$ satisfies Eq. (2.5).*

Proof. The last lemma and the GP equation imply $\nabla^2 \Phi^{\text{GP}} \in L^\infty_{\text{loc}}$. Thus $\nabla \Phi^{\text{GP}}$ exists and is Hölder continuous (see [9], Sec. 10.2). The C^∞ property follows by a bootstrap argument. The differentiability with respect to the parameter a may be shown by a Feynman-Hellmann type argument like analogous statements (e.g., with respect to differentiability nuclear charges) in Thomas-Fermi theory [8]. Equation (2.5) follows immediately from Eq. (1.6) and $E^{\text{GP}}(1, Na) = \mathcal{E}^{\text{GP}}[\Phi^{\text{GP}}_{1, Na}]$. $\square$

Lemmas A.1–A.6 complete the proof of Theorem II.1.

APPENDIX B

In this appendix we show that Eq. (4.1) holds for non-negative potentials satisfying Eq. (4.2), and that a similar estimate with $1/17$ replaced by $O(\varepsilon)$ holds under the condition (4.3).

We cut the potential at a finite radius $\bar{R}$ which, because of $v \geq 0$, can only decrease the energy. We thus define

$$\tilde{v}(r) = v(r) \Theta(\bar{R} - r) \quad (\text{B1})$$

and denote the corresponding scattering length by $\tilde{a} \leq a$. Let u be the zero-energy scattering solution for the potential v [cf. Eq. (1.2)] and put

$$h(r) = r - \frac{u(r)}{u'(r)}. \quad (\text{B2})$$

The difference $a - \tilde{a}$ can be estimated as follows. Since $v(r)$ and $\tilde{v}(r)$ agree for $r \leq \bar{R}$, the same holds for the corresponding scattering solutions. Moreover, $\tilde{a} = h(\bar{R})$. Hence

$$\begin{aligned} a - \tilde{a} &= \int_{\bar{R}}^{\infty} h'(r) dr = \int_{\bar{R}}^{\infty} \frac{u(r)u''(r)}{u'(r)^2} dr \\ &\leq \int_{\bar{R}}^{\infty} \frac{u''(r)}{u(r)} r^2 dr = \frac{1}{2} \int_{\bar{R}}^{\infty} v(r) r^2 dr, \end{aligned} \quad (\text{B3})$$

where convexity of u has been used to derive the inequality. We remark that for $\bar{R} \rightarrow 0$ this simple estimate gives the *Spruch-Rosenberg* inequality [12]

$$a \leq \frac{1}{2} \int_0^{\infty} v(r) r^2 dr. \quad (\text{B4})$$

Assuming Eq. (4.2) one obtains

$$\tilde{a} \geq a \left(1 - \text{const} \left(\frac{a}{\bar{R}} \right)^{1/5 + \epsilon} \right). \quad (\text{B5})$$

Equation (4.1) holds in any case with a replaced by $\tilde{a}$, and if we take $\bar{R} \propto a Y^{-5/17 + \epsilon'}$ with $\epsilon' > 0$ then the error in Eq. (B5) is of higher order than the leading error term in Eq. (4.1). We have thus established Eq. (4.1) under the condition (4.2). If only the weaker condition (4.3) holds, then the additional error term may be $O(Y^{(5/17 - \epsilon')\epsilon})$.

To see the significance of condition (4.3) we also estimate $a - \tilde{a}$ from below:

$$a - \tilde{a} \geq \int_{\bar{R}}^{\infty} u(r) u''(r) dr \geq \frac{1}{2} \int_{\max(\bar{R}, a)}^{\infty} v(r) (r - a)^2 dr. \quad (\text{B6})$$

In order that a be finite the last integral must converge, i.e., a slower decrease than $1/r^3$ is not allowed.

- [1] E.P. Gross, *Nuovo Cimento* **20**, 454 (1961).
- [2] L.P. Pitaevskii, *Zh. Eksp. Teor. Fiz.* **40**, 646 (1961) [*Sov. Phys. JETP* **13**, 451 (1961)].
- [3] E.P. Gross, *J. Math. Phys.* **4**, 195 (1963).
- [4] F. Dalfovo, S. Giorgini, L.P. Pitaevskii, and S. Stringari, *Rev. Mod. Phys.* **71**, 463 (1999).
- [5] E.H. Lieb and J. Yngvason, *Phys. Rev. Lett.* **80**, 2504 (1998).
- [6] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV* (Academic Press, New York, 1978).

- [7] F.J. Dyson, *Phys. Rev.* **106**, 20 (1957).
- [8] E.H. Lieb, *Rev. Mod. Phys.* **53**, 603 (1981).
- [9] E. H. Lieb and M. Loss, *Analysis* (American Mathematical Society, Providence, RI, 1997).
- [10] P.L. Lions, *Appl. Anal.* **12**, 267 (1981).
- [11] E.H. Lieb, *Phys. Rev.* **130**, 2518 (1963); see also E.H. Lieb and A.Y. Sakakura, *Phys. Rev. A* **133**, A899 (1964); E.H. Lieb and W. Liniger, *ibid.* **134**, A312 (1964).
- [12] L. Spruch and L. Rosenberg, *Phys. Rev.* **116**, 1034 (1959).

THE $N^{5/3}$ LAW FOR BOSONSElliott H. LIEB¹*Departments of Mathematics and Physics, Princeton University, Princeton, NJ 08540, USA*

Received 13 December 1978

Non-relativistic negative bosons interacting with infinite mass positive particles via Coulomb forces are shown to be unstable in the sense that $E_0 \leq -CN^{5/3}$. This agrees with the previously known lower bound $E_0 \geq -AN^{5/3}$.

In a celebrated series of papers [1–3], Dyson and Lenard proved that matter is quantum-mechanically stable under the action of Coulomb forces provided all the particles of at least one sign of charge (say negative) are fermions. In other words, the ground state energy E_0 satisfies

$$E_0 \geq -A_f N \quad (\text{negative fermions}), \quad (1)$$

where N is the number of negative particles. It is not necessary to assume neutrality or that the positive particles have finite mass. It is necessary to assume that all the charges are bounded however. (1) was subsequently rederived by Federbush [4] and by Lieb and Thirring [5]. The best current value [6] is

$$A_f \leq -22.24 \text{ Ry}, \quad (2)$$

for electrons and protons, and with the assumption of neutrality.

If all the particles are bosons or, what is the same thing, are not subject to any statistics, the best available lower bound [1–7] is

$$E_0 \geq -A_b N^{5/3} \quad (\text{all bosons}), \quad (3)$$

with [6]

$$A_b \leq 14.01 \text{ Ry}, \quad (4)$$

when the positive and negative particles have charges $\pm e$ and the negative particle mass is m_e . The positive particles can have infinite mass.

Dyson [8] then proved, by a complicated variational calculation, that in the boson case

$$E_0 \leq -BN^{7/5}, \quad (5)$$

if all particles have finite mass. It was conjectured [1,8] that $N^{7/5}$ is the correct law for bosons and not $N^{5/3}$. While this question might have only moderate practical importance (it would be relevant for π^- mesons and ^4He nuclei, for example), it has great theoretical importance and it is to be hoped that its solution will soon be forthcoming. It is interesting because at this point there is no simple, compelling physical argument, as distinguished from a computational argument [9], why the $N^{7/5}$ law is correct. Subtle correlation effects are yet to be understood fully and rigorously.

The purpose of this paper is to add a minor commentary on the problem. By means of a simple variational calculation, it will be shown that the $N^{5/3}$ law is indeed correct if the positive particles have *infinite mass* and charge $z|e| > 0$, i.e.

$$E_0 \leq -CN^{5/3}. \quad (6a)$$

If the negative particles have mass m_e and charge $-|e|$, and if the system is neutral, then

$$C \geq (1/108)z^{4/3} \text{ Ry}. \quad (6b)$$

Thus, the limits $N \rightarrow \infty$ and the mass of the positive particles $\rightarrow \infty$ are not interchangeable if the $N^{7/5}$ conjecture is correct.

We note in passing that (1) alone does not imply that $e_0 = \lim_{N \rightarrow \infty} E_0/N$ exists. However, the method

¹ Work partially supported by U.S. National Science Foundation grants MCS 75-21684 A02 and INT 78-01160.

developed to prove the existence of the thermodynamic limit [10] also proves that this limit exists. Likewise (3) and (6a) do not imply that $\lim_{N \rightarrow \infty} E_0/N^{5/3}$ exists. The aforementioned method [10] is not suitable for this task and we do not know an adequate substitute.

For simplicity of exposition we assume there is one kind each of positive and negative particle. The N negative particles have a mass $=1/2$ and charge -1 . If $\hbar^2 = 1$, then one Rydberg $= 1/4$. The K infinite mass positive particles have charge $z > 0$ and we assume $N = Kz$ (neutrality) with $K = 8n^3$, n an integer. Other cases can easily be handled by this method but we omit details for simplicity. The hamiltonian for the negative particles is

$$H_{N,R} = - \sum_{i=1}^N \{ \Delta_i + V_R(r_i) \} + \sum_{1 \leq i < j \leq N} |r_i - r_j|^{-1} + U(R), \quad (7)$$

where $R = \{R_1, \dots, R_K\}$ is the collection of fixed coordinates of the positive particles and

$$V_R(r) = z \sum_{j=1}^K |r - R_j|^{-1}, \quad U(R) = z^2 \sum_{1 \leq i < j \leq K} |R_i - R_j|^{-1}. \quad (8)$$

We want to find a normalized $\psi(r_1, \dots, r_N)$ and R (depending on N) such that $\langle \psi | H_{N,R} | \psi \rangle \leq -CN^{5/3}$.

ψ is chosen to be a simple product of identical functions:

$$\psi(r_1, \dots, r_N) = \prod_{i=1}^N \phi_\lambda(r_i). \quad (9)$$

ϕ_λ depends on the parameter λ (which will turn out to be proportional to $N^{1/3}$) as follows:

$$\phi_\lambda(r) = \lambda^{3/2} g(\lambda r), \quad (10)$$

where $g(r)$ is the fixed, normalized function

$$g(r) = f(x)f(y)f(z), \quad r = (x, y, z), \quad (11)$$

and

$$f(x) = \sqrt{3/2} [1 - |x|], \quad |x| \leq 1, \\ = 0, \quad |x| \geq 1. \quad (12)$$

The explicit choice in (12) is neither important nor optimal.

With

$$T = \int |\nabla g(r)|^2 d^3r = 9, \quad (13)$$

$$\langle \psi | H_{N,R} | \psi \rangle \leq \lambda^2 NT + \lambda W(N, R), \quad (14)$$

$$W(N, R) = \frac{1}{2} N^2 \iint g^2(r) g^2(r') |r - r'|^{-1} d^3r d^3r' - N \int g^2(r) V_R(r) d^3r + U(R). \quad (15)$$

There is $\leq$ in (14) because we should have $N(N-1)$ instead of N^2 in (15). We claim R can be chosen so that

$$W(N, R) \leq -(12)^{-1/2} z^{2/3} N^{4/3}. \quad (16)$$

If so, (6) is proved by minimizing (14) with respect to λ .

To prove (16), let $0 = a(0) < a(1) < \dots < a(n) = 1$ (recall $K = 8n^3$) and define the real intervals $L(j) = [a(j), a(j+1)]$ for $0 \leq j \leq n-1$ and $L(j) = [-a(-j), -a(-j-1)]$ for $-n \leq j \leq -1$. Choose the $a(j)$ such that

$$\int_{L(j)} f(x)^2 dx = (2n)^{-1}, \quad \text{for all } j. \quad (17)$$

The rectangles $\Gamma(i, j, k) = L(i) \times L(j) \times L(k)$, $-n \leq i, j, k < n$ satisfy

$$\int_{\Gamma(i,j,k)} g(r)^2 d^3r = 1/K. \quad (18)$$

Returning to (15), place one R_i in each of the K rectangles and then average $W(N, R)$ over the positions, R_i , of the positive particles within the rectangles with a relative weight $g(R_1)^2 \dots g(R_K)^2$. This average is given by

$$W(N) = -\frac{1}{2} N^2 \sum_m \int_{\Gamma(m)} g(r)^2 g(r')^2 \times |r - r'|^{-1} d^3r d^3r', \quad (19)$$

with $m = (i, j, k)$. Since the weight is nonnegative, there is at least one choice of R (with one particle per

rectangle) such that $W(N, R) \leq W(N)$. Thus, we are done if $W(N) \leq$ right side of (16).

Each integral in (19) is the self-energy of a charge $1/K$ confined to a rectangle. This, in turn, is greater than the *minimum* self-energy of a charge $1/K$ confined to lie only in a circumscribed sphere. If $\Gamma(m)$ has sides of length (s, t, u) , this sphere has radius $\rho(m) = \frac{1}{2} [s^2 + t^2 + u^2]^{1/2}$. It is well known that the minimum self-energy occurs when the charge is distributed uniformly on the surface of the sphere and is $\rho(m)^{-1} \times (1/K)^2$. Thus

$$\begin{aligned} W(N) &\leq -\frac{1}{2} z^2 \sum_m \rho(m)^{-1} \\ &\leq -z^2 K [\sigma^2 + \tau^2 + \mu^2]^{-1/2} \quad (\text{by convexity}), \end{aligned} \quad (20)$$

where σ, τ and μ are the mean lengths of the sides of the rectangles. However, $\sigma = \tau = \mu = 1/n$, and (16) is proved.

The author would like to thank the Research Institute for Mathematical Sciences, Kyoto University, in particular Professors H. Araki and K. Itô, for their generous hospitality.

References

- [1] F.J. Dyson and A. Lenard, J. Math. Phys. 8 (1967) 423; A. Lenard and F.J. Dyson, J. Math. Phys. 9 (1968) 698.
- [2] F.J. Dyson, in: Brandeis University Summer Institute in Theoretical physics (1966) eds. M. Chretien, E.P. Gross and S. Deser (Gordon and Breach, New York 1968), Vol. 1, p. 179.
- [3] A. Lenard, in: Statistical mechanics and mathematical problems, ed. A. Lenard, Lecture Notes in Physics (Springer, New York, 1973), Vol. 20, p. 114.
- [4] P. Federbush, J. Math. Phys. 16 (1975) 347, 706.
- [5] E.H. Lieb and W.E. Thirring, Phys. Rev. Lett. 35 (1975) 687.
- [6] E.H. Lieb, Rev. Mod. Phys. 48 (1976) 553.
- [7] D. Brydges and P. Federbush, J. Math. Phys. 17 (1976) 2133.
- [8] F.J. Dyson, J. Math. Phys. 8 (1967) 1538.
- [9] L.L. Foldy, Phys. Rev. 124 (1961) 649; M. Girardeau and R. Arnowitt, Phys. Rev. 113 (1959) 755; M. Girardeau, Phys. Rev. 127 (1962) 1809; W.H. Bassichis and L.L. Foldy, Phys. Rev. 133 (1964) A935; W.H. Bassichis, Phys. Rev. 134 (1964) A543; J.M. Stephen, Proc. Phys. Soc. (London) 79 (1962) 994; D.K. Lee and E. Feenberg, Phys. Rev. 137 (1965) A731; D. Wright, Phys. Rev. 143 (1966) 91; E.H. Lieb, Phys. Rev. 130 (1963) 2518; E.H. Lieb and A.Y. Sakakura, Phys. Rev. 133 (1964) A899.
- [10] E.H. Lieb and J.L. Lebowitz, Adv. Math. 9 (1972) 316.

The $N^{7/5}$ Law for Charged Bosons

Joseph G. Conlon^{1*}, Elliott H. Lieb^{2**} and Horng-Tzer Yau^{2***}

¹ Department of Mathematics, University of Missouri, Columbia, MO 65211, USA

² Departments of Mathematics and Physics, Princeton University, P.O.B. 708, Princeton, NJ 08544, USA

Abstract. Non-relativistic bosons interacting with Coulomb forces are unstable, as Dyson showed 20 years ago, in the sense that the ground state energy satisfies $E_0 \leq -AN^{7/5}$. We prove that $7/5$ is the correct power by proving that $E_0 \geq -BN^{7/5}$. For the non-relativistic bosonic, one-component jellium problem, Foldy and Girardeau showed that $E_0 \leq -CN\rho^{1/4}$. This $1/4$ law is also validated here by showing that $E_0 \geq -DN\rho^{1/4}$. These bounds prove that the Bogoliubov type paired wave function correctly predicts the order of magnitude of the energy.

I. Introduction and Background

Twenty years ago Dyson and Lenard [5] proved the stability of ordinary non-relativistic matter with Coulomb forces, namely that the ground state energy, E_0 , of an N -particle system satisfies $E_0 \geq -A_1N$ for some universal constant A_1 . In ordinary matter, the negative particles (electrons) are fermions. At the same time, Dyson [4] proved that bosonic matter is definitely not stable; if all the particles (positive as well as negative) are bosons then $E_0 \leq -A_2N^{7/5}$ for some $A_2 > 0$. Dyson and Lenard [5] did prove, however, that $E_0 \geq -A_3N^{5/3}$ in the boson case, and thus the open problem was whether the correct exponent for bosons is $5/3$ or $7/5$ or something in between.

In this paper we prove that the $N^{7/5}$ law is correct for bosons by obtaining a lower bound $E_0 \geq -A_4N^{7/5}$. As is well known, the bosonic energy is the absolute lowest energy when no symmetry restriction is imposed.

It may appear that the difference between $5/3$ and $7/5$ is insignificant, especially since bosonic matter does not exist experimentally, but that impression would fail to take into account the essential difference between the ground states implied by

* Work partially supported by U.S. National Science Foundation grant DMS 8600748

** Work partially supported by U.S. National Science Foundation grant PHY85-15288-A01

***Work supported by Alfred Sloan Foundation dissertation fellowship

the two laws. The $-A_3 N^{5/3}$ lower bound can be derived by using a semiclassical estimate which leads to a Thomas–Fermi type theory. This estimate is the same as that used by Lieb and Thirring [15] to give a simple proof of the stability of matter in the fermion case. Correlations are unimportant in this estimate. The $N^{7/5}$ law, on the other hand, is much more subtle. To get the upper bound, $-A_2 N^{7/5}$, Dyson had to use an extremely complicated variational function which contains delicate correlations. It is the same kind of function proposed by Bogoliubov [1] (see also [10] for a review) in his theory of the many-boson system and in which particles of equal and opposite momenta are paired. It is also very similar to the Bardeen–Cooper–Schrieffer pair function of superconductivity. Since this kind of wave function plays such an important role in physics, it is important to know whether it is correct, and in proving the $N^{7/5}$ law for the energy we are, in a certain sense, validating this function.

The Hamiltonian to be considered is

$$H_N = - \sum_{i=1}^N \Delta_i + \sum_{1 \leq i < j \leq N} e_i e_j |x_i - x_j|^{-1}, \quad (1.1)$$

which is relevant for N charged particles with coordinates labeled $x_1, \dots, x_N \in \mathbf{R}^3$. The charges satisfy $e_i = \pm 1$, all i and $\hbar^2/2m = 1$. The neutral case is $\sum e_i = 0$. In Sect. II we shall prove the $N^{7/5}$ lower bound for H_N which is stated precisely in Theorem 1.2 below. The neutrality condition is not imposed in this theorem. If, however, the system is very non-neutral, with N_- negative and N_+ positive particles with $N_- + N_+ = N$ and $N_+ \gg N_-$, we expect that the bounds (1.7) and (1.8) are not optimal. One should have $E_0 \geq -A_5 N^{7/5}$ instead; this is indeed true but, for simplicity of exposition, this generalization is deferred to Sect. V, Theorem 5.1.

A closely related system that we shall consider in Sect. III is jellium. In this case there is a domain Λ , in which there is a fixed constant density, ρ_B , of positive charge called the background. There are also N negative particles of charge -1 and the jellium Hamiltonian is

$$H_{N,\Lambda}' = - \sum_{i=1}^N \{\Delta_i + V(x_i)\} + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} + \frac{1}{2} \rho_B \int_{\Lambda} V(x) dx, \quad (1.2)$$

where $V(x) = \rho_B \int_{\Lambda} |x - y|^{-1} dy$ is the potential generated by the background. We do not restrict ourselves to the neutral case, $N = \rho_B L^3$, in Sect. III. As boundary conditions we can take either $\psi \in L^2(\mathbf{R}^{3N})$ or else $\psi \in L^2(\Lambda^N)$ with Dirichlet Boundary conditions. Clearly E_0 for the former is less than E_0 for the latter. In the physics literature one usually imposes neutrality and takes Λ to be a cube, $\psi \in L^2(\Lambda^N)$ with periodic boundary conditions and, in addition, the potential is replaced by an interaction solely among the negative particles in which the $k=0$ Fourier component of the $1/r$ potential is omitted. It is not a trivial matter to show rigorously that this periodic problem is the same, in the thermodynamic limit, as the more physical problem (1.2) which we consider here—even in the neutral case. Here, again, the bosonic energy is the absolute lowest.

Let us briefly review what is known rigorously about these two problems.

A. Jellium. Foldy [7] was the first to apply Bogoliubov's method to the neutral

bosonic jellium problem (with the periodic boundary conditions mentioned above) and obtained, for large ρ_B and in the thermodynamic limit,

$$E_0 \approx -1.933N\rho_B^{1/4}. \quad (1.3)$$

A proof of (1.3) was, and is lacking, but later, Girardeau [8] proved that (1.3) is an upper bound to E_0 (for large ρ_B and with the same conditions). Another non-rigorous derivation of (1.3) that does not use Bogoliubov's method was given by Lieb and Sakakura [13]. In Sect. III we shall derive the following lower bound for the real problem (1.2).

Theorem 1.1. *With $H_{N,\Lambda}^J$ given by (1.2) on $L^2(\mathbf{R}^{3N})$, the ground state energy satisfies, for all N and Λ*

$$E_0 \geq -A_6 N \rho_B^{1/4}, \quad (1.4)$$

for some universal constant A_6 . A bound for A_6 is given in (3.19); In the limit $\rho_B \rightarrow \infty$ we can take $A_6 = 8.57$.

Theorem 1.1 is generalized in Theorem 3.1. Note that our lower bound (1.4) is close to the upper bound (1.3) (with a factor about 4.5).

The existence of the thermodynamic limit for jellium was proved by Lieb and Narnhofer [12]. This limit will not concern us here, but a useful result in the appendix of [12] contains a lower bound to the potential energy terms in (1.2), and hence to the ground state energy of (1.2) for all N . This bound is

$$E_0 \geq -(0.9)(4\pi/3)^{1/3} N \rho_B^{1/3}. \quad (1.5)$$

A result similar to (1.5) is given in [3].

It is not easy to give a heuristic derivation of the $\rho_B^{1/4}$ law. Dyson [4] gives one, but we prefer the following point of view. The reason that $E_0 < 0$ is that the negative particles stay away from each other. If λ is the correlation length (i.e. the radius of a ball surrounding any one particle in which there is, on the average, an absence of one particle) then the potential energy, P , is roughly $P \approx -N/\lambda$. On the other hand, let us study the kinetic energy, K . Most of the particles will be in the zero momentum state. A correlation length λ can be achieved by decomposing Λ into $n \equiv (L/\lambda)^3$ boxes of size λ . If there is one single particle wave function in each box, with Dirichlet boundary conditions, its kinetic energy will be λ^{-2} and thus $K = n\lambda^{-2} = L^3\lambda^{-5}$. Minimization of $K + P = -N\lambda^{-1} + L^3\lambda^{-5}$ with respect to λ (recalling $N = \rho_B L^3$ for neutrality) yields $\lambda^4 = 5/\rho_B$ and $E_0 = -\frac{4}{5}N\lambda^{-1} = -\frac{4}{5}N(\rho_B/5)^{1/4}$. In addition we learn that $K/P = -1/5$, which is very different from the usual virial theorem value $-1/2$.

The difficulty with the above argument is its apparent inconsistency. If we put n particles into boxes, as stated above, then K will be $n\lambda^{-2}$ but also P will be $-n\lambda^{-1}$, not $N\lambda^{-1}$. Nevertheless, it is true that the Bogoliubov pair wave function has the properties $K \approx n\lambda^{-2}$ and $P = -N\lambda^{-1}$ mentioned above. How it achieves this is not easy to understand; one *must*, apparently, study the problem in momentum space.

If the kinetic energy were $|p|^\alpha$ with $1 \leq \alpha < 2$, instead of p^2 , we would predict, by the same argument, that E_0 would then be of the order $-N\rho_B^{1/(2+\alpha)}$ and $\lambda \approx \rho_B^{-1/(2+\alpha)}$. This conclusion does indeed agree with what is obtained from an

appropriately modified Bogoliubov function. When $\alpha = 1$ (the relativistic case) we get $-N\rho_B^{1/3}$ which agrees with the lower bound (1.5).

B. The Two-Component System. For simplicity let us consider the neutral system with N bosons of each charge. In [5] and [15] it is proved that $E_0 \geq -A_3 N^{5/3}$. Indeed, if one kind of particle is infinitely massive then the $N^{5/3}$ law is correct—as proved by Lieb [9]. Moreover, the $N^{5/3}$ upper bound in [9] is very simple and semiclassical—correlations are unnecessary.

The $N^{7/5}$ result for particles all of finite mass is subtle. For (1.1) Dyson obtained (for large N and $\sum e_i = 0$)

$$E_0 \leq -5.001 \times 10^{-7} N^{7/5}. \tag{1.6}$$

Surely, the coefficient in (1.6) is too small. Our lower bound for the energy, proved in Sect. II, is the following:

Theorem 1.2. *Let H_N be given by (1.1) with $e_i = \pm 1$. Neutrality is not assumed. Then, on $L^2(\mathbf{R}^{3N})$*

$$H_N \geq -0.30 N^{7/5} \tag{1.7}$$

for sufficiently large N .

Generalizations of Theorem 1.2 are given in Theorems 2.1 and 5.1. The former is a generalization to the Yukawa potential while the latter treats the nonneutral case $N_- \ll N_+$,

$$E_0 \geq -A_7 N^{7/5} \tag{1.8}$$

for some constant A_7 .

Let us recall Dyson's heuristic derivation [4] of (1.7) from (1.4). There are two parts to the energy: (i) a local kinetic energy and electrostatic correlation energy and (ii) a global kinetic energy needed to localize the system in a region of radius R . The latter is approximately $K_{\text{global}} \approx N/R^2$. The former is approximately $E_{\text{local}} \approx -A_5 N \rho^{1/4}$ with $\rho = N/R^3$. Here we have taken over the one-component jellium result (1.4) even though we are considering a two-component system; the reason is that the electrostatic correlation energy comes primarily from the fact, as we said, that particles of like charge stay away from each other and therefore the energy in the two-component and one-component systems are comparable. If we now minimize $E = E_{\text{global}} + E_{\text{local}}$ with respect to R we find $R \approx N^{-1/5}$ and $E \approx -N^{7/5}$. A check on the consistency of this is that the correlation length satisfies $\lambda \approx \rho^{-1/4} = (N/R^3)^{-1/4} = N^{-2/5} \ll R$.

In the present paper we begin with the $N^{7/5}$ problem and prove (1.7) in Sect. II. Our analysis is based on Conlon's paper [2] in which the following was proved about the two-component system in a box. A symmetric wave function connotes a function that is separately symmetric in the positive and negative charge spatial variables, i.e. a bosonic function.

Theorem 1.3. *Let Λ be a cube in $\mathbf{R}^3$ and suppose that $\psi(x_1, \dots, x_N)$ is any symmetric, infinitely differentiable, $L^2(\mathbf{R}^{3N})$ normalized function with support in Λ^N . Let*

$$K(\psi) \equiv \sum_{i=1}^N \langle \psi, -\Delta_i \psi \rangle \equiv \langle \psi, T\psi \rangle \tag{1.9}$$

be the kinetic energy and define γ_ψ by

$$K(\psi) = N\gamma_\psi^2/L^2, \quad (1.10)$$

where L is the length of Λ . Let H_N^ν be the two-component Hamiltonian analogous to (1.1) but with the Coulomb potential replaced by the Yukawa potential $Y_\nu(x) = |x|^{-1} \exp(-\nu|x|)$, namely

$$H_N^\nu = - \sum_{i=1}^N \Delta_i + \sum_{1 \leq i < j \leq N} e_i e_j Y_\nu(x_i - x_j).$$

The $e_i = \pm 1$ as before and neutrality is not assumed. Then, if $\gamma_\psi \leq N^{1/3}$,

$$\langle \psi, H_N^\nu \psi \rangle \geq -A_8 N^{7/5} \quad (1.11)$$

for some constant A_8 , which is independent of ν , N and L .

Theorem 1.3 is proved in [2] by a succession of inequalities that turn the Bogoliubov ansatz [1] into a rigorous bound (with a different constant, of course). It concerns the local energy and, being intrinsically quantum-mechanical, has no classical, analogue. The reason that Theorem 1.3 does not imply the $N^{7/5}$ law, Theorem 1.2, is the condition that $\gamma_\psi \leq N^{1/3}$ (alternatively, $K(\psi) \leq N^{5/3}/L^2$). We do not know in advance what the radius, R , is for an energy minimizer. If, for example, $K(\psi) = N^{5/3}$ and $R \gg 1$, we could not use Theorem 1.3. Thus, we are faced with what might be called an infrared problem which our analysis in Sect. II solves.

To get the constant in (1.7) we need a good value for A_8 in (1.11). A value can be deduced from [2], but the constant there is not optimum. It turns out that restricting $\gamma_\psi \leq N^{1/3-\delta}$ for some $\delta > 0$ is sufficient for the analysis in Sect. II. Under this condition the following improvement of Theorem 1.3 is possible, and is proved in Sect. IV.

Theorem 1.4. Assume the hypotheses of Theorem 1.3 except that $\gamma_\psi \leq N^{1/3}$ is replaced by $\gamma_\psi \leq N^{1/3-\delta}$ for some fixed $\delta > 0$. The parameters ν and L can depend on N , but we assume that $N^{-1/5}\nu L$ stays bounded as $N \rightarrow \infty$. Then, for sufficiently large N ,

$$\langle \psi, H_N^\nu \psi \rangle \geq -0.30 N^{7/5}. \quad (1.12)$$

The analysis in Sect. III of the jellium problem, leading to (1.4), uses the $N^{7/5}$ result of Sect. II. This may seem a bit odd in view of Dyson's heuristic discussion in which one uses the jellium result to understand the $N^{7/5}$ theorem. Our procedure is to bound the jellium energy in arbitrarily large boxes in terms of the energy in a box of size $l = \rho_B^{-1/8}$. In such a box the particle number (with neutrality) is $n = \rho_B l^3 = \rho_B^{5/8}$. But then, by the $N^{7/5}$ theorem (with the background being thought of as N particles in a simple, smeared out state), $E_{\text{box}} \geq -An^{7/5} = -A\rho_B^{7/8} = -An\rho_B^{1/4}$. By adding up the boxes we obtain $E_0 \geq -AN\rho_B^{1/4}$.

Our work here leads to many questions, of which the following are a few.

Open Problems

- (1) Find the correct coefficient in (1.4) for large ρ_B in the jellium problem. Is Foldy's constant in (1.3) correct?
- (2) Find the correct coefficient A_7 in (1.8) as $N \rightarrow \infty$ for the two-component

problem. The bound in Theorem 1.4 is within a factor of 11 of what one would get heuristically from a calculation using the Bogoliubov function. This is discussed in Sect. IV. This bound translates into the bound (1.7) of Theorem 1.2. We should emphasize, however, that Bogoliubov's method does not predict an exact value for the asymptotic constant in Theorem 1.2. The reason for this is that in the Bogoliubov method one is forced to work in cubes and, in the Bogoliubov function, most particles are in the lowest momentum mode of the cube. The size of the cube can be taken to be the size of the system, namely $N^{-1/5}$. Thus the energy depends critically on the lowest eigenvalue of the Laplacian in a cube and this depends on boundary conditions. The lowest eigenvalue will be uncertain because of the boundary conditions and will be of order $N^{2/5}$. The uncertainty in the energy will be of order $N^{7/5}$.

(3) What can be said about the correlation functions at high density? Is Bogoliubov's ansatz really correct or does it merely give a good account of the energy?

(4) As shown in Sect. II, the statement $E_0 \geq -AN^{7/5}$ for the two-component system is equivalent, via the virial theorem, to $-P(\psi) \leq 2A^{1/2}N^{7/10}K(\psi)^{1/2}$ for all ψ . Here $K(\psi)$ is the kinetic energy and $P(\psi)$ the potential energy of ψ . Now let us replace p^2 by $|p|^\alpha$ in the kinetic energy. In the heuristic discussion above we surmised that the jellium energy should be $-C_\alpha \rho_B^{1/(2+\alpha)}$. Then, by the uncertainty principle argument relating the jellium energy to the two-component energy given before, we would have (with $K_{\text{global}} \approx NR^{-\alpha}$) $R \approx N^{-1/(\alpha-1)(\alpha+3)}$ and $E_0 = -A_\alpha N^{(\alpha^2+3\alpha-3)/(\alpha^2+2\alpha-3)}$. This statement about E_0 is equivalent, via the virial theorem, to

$$-P(\psi) \leq \alpha[A_\alpha/(\alpha-1)]^{1-1/\alpha}N^{(\alpha^2+3\alpha-3)/(\alpha(\alpha+3))}K(\psi)^{1/\alpha}. \tag{1.13}$$

We conjecture that these formal calculations are correct as $N \rightarrow \infty$. If so, it is interesting to look at the $\alpha = 1$ case (relativistic bosons). In this case, $E_0 = -\infty$ for large enough N , which is correct, but (1.13) continues to make sense. Namely, for $\alpha = 1$,

$$-P(\psi) \leq CN^{1/4}K(\psi). \tag{1.14}$$

We conjecture that (1.14) is true for large N and we remark that in [3] it is proved that (1.14) holds with $N^{1/4}$ replaced by $N^{1/3}$. Since the bosonic energy is the absolute lowest, (1.13) and (1.14) are independent of statistics.

II. The $N^{7/5}$ Theorem

Our strategy to prove Theorem 1.2 is to decompose $\mathbf{R}^3$ into cubic boxes of size $l = N^{-\varepsilon}$ with ε some small number less than $1/5$. This l is large compared to the expected size of the system, $N^{-1/5}$, but we do not know this fact in advance. It will be necessary to localize H_N in these boxes and to control the interaction between boxes.

The main difficulty in localizing the Hamiltonian (1.1) comes from the localization of the Coulomb potential. The effects of localization on the kinetic energy can be easily computed to be of order Nl^{-2} , where l is the cutoff length. For the potential energy, however, even a small amount of net charge will produce

enormous potential energies, and therefore charge neutrality must be preserved very carefully. Our basic strategy is first to replace the Coulomb potential by a Yukawa cutoff and then, by averaging over all possible box locations, the errors can be controlled.

For $\mu > 0$, let

$$Y_\mu(r) \equiv r^{-1} e^{-\mu r} \quad (2.1)$$

be the Yukawa potential with range μ^{-1} . For $\kappa > 0$ and N a positive integer, let

$$H_{\kappa N}^{\mu l} \equiv - \sum_{i=1}^N \Delta_i + \kappa \sum_{1 \leq i < j \leq N} e_i e_j Y_\mu(x_i - x_j) \quad (2.2)$$

be the Hamiltonian of N charged bosons interacting pairwise by the Yukawa potential with coupling constant κ . As before, $e_i = \pm 1$ but neutrality is not assumed. $H_{\kappa N}^{\mu l}$ is defined as a quadratic form on $L^2([0, l]^{3N})$ with Dirichlet boundary conditions. We shall drop μ or l or κ whenever they are equal to 0, ∞ or 1 respectively. Since the Hamiltonian (2.2) is symmetric under permutations separately on positive or negative particles, the ground state automatically obeys Bose statistics, and we shall assume this henceforth. Let

$$E_{\kappa N}^{\mu l} = \inf \text{spec } H_{\kappa N}^{\mu l} \quad (2.3)$$

be the ground state energy. Then a trivial scaling yields, for the Coulomb Hamiltonian in all of $\mathbf{R}^3$,

$$E_{\kappa N} = \kappa^2 E_N. \quad (2.4)$$

To fix a partition, let g be a piecewise C^1 function on $\mathbf{R}$ defined by

$$g(t) = \cos(\pi t/2), \quad -1 \leq t \leq 1 \quad (2.5)$$

and zero otherwise. Then $\sum_{j \in \mathbf{Z}} g^2(t+j) = 1$ for all $t \in \mathbf{R}$. Let $\chi(x) \equiv g(x^1)g(x^2)g(x^3)$, with $x = (x^1, x^2, x^3)$ and let $\chi_{u\lambda}(x) \equiv \chi(x+u+\lambda)$. Here $\lambda \in \mathbf{Z}^3$ and $u \in [0, 1]^3 \equiv \Gamma$. Then

$$\sum_{\lambda \in \mathbf{Z}^3} \chi_{u\lambda}^2(x) = 1 \quad \forall x \in \mathbf{R}^3, \quad u \in \Gamma. \quad (2.6)$$

A function h which is of central importance in our localization is defined by

$$h(x, y) \equiv \int_{\Gamma} du \sum_{\lambda \in \mathbf{Z}^3} \chi_{u\lambda}^2(x) \chi_{u\lambda}^2(y). \quad (2.7)$$

Then h depends only on the difference $z \equiv x - y$ and

$$\begin{aligned} h(z) &\equiv h(x - y) = \int_{\Gamma} du \sum_{\lambda \in \mathbf{Z}^3} \chi^2(x + u + \lambda) \chi^2(y + u + \lambda) \\ &= \int_{\mathbf{R}^3} du \chi^2(x + u) \chi^2(y + u) = (\chi^2 * \chi^2)(z). \end{aligned} \quad (2.8)$$

An easy computation shows that

$$\begin{aligned} (g^2 * g^2)(t) &= \frac{1}{8} \left[4 - 2|t| + \frac{3}{\pi} \sin \pi |t| + (2 - |t|) \cos \pi t \right], \quad |t| \leq 1 \\ &\sim \frac{3}{4} - \frac{1}{8} \pi^2 t^2 + O(t^4) \end{aligned} \quad (2.9)$$

and zero otherwise. Hence $h(z)$ is a C^4 function and

$$|h(z) - a_0 - a_1|z|^2| \leq a_2|z|^4 \tag{2.10}$$

with $a_0 = (3/4)^3, a_1 = -(\frac{3}{4})^2(\pi^2/8)$ and a_2 some constant of order 1.

Let $h_l(x) \equiv h(x/l)$.

We now define localized kinetic and potential energies. Let $\alpha = (u, \alpha_1, \dots, \alpha_N) \in \Gamma \times \mathbb{Z}^{3N}$ be a multi-index and let $\int d\alpha \equiv \int du \sum_{\alpha_1 \in \mathbb{Z}^3} \dots \sum_{\alpha_N \in \mathbb{Z}^3}$. If $\beta = (v, \beta_1, \dots, \beta_N)$ is another multi-index, denote $\delta(u-v) \prod_{i=1}^N \delta_{\alpha_i \beta_i}$ by $\delta_{\alpha \beta}$. (Here, $\delta(u-v)$ is the Dirac δ -function and $\delta_{\alpha_i \beta_i}$ is the Kronecker delta.) For any $l > 0$, let

$$\psi^l_\alpha(x_1, \dots, x_N) \equiv \prod_{k=1}^N \chi_{u\alpha_k}(x_k/l) \psi(x_1, \dots, x_N), \tag{2.11}$$

$$V^\mu_\alpha \equiv \sum_{1 \leq i < j \leq N} e_i e_j Y_\mu(x_i - x_j) \delta_{\alpha_i, \alpha_j}. \tag{2.12}$$

Then by using (2.6) and (2.7) one has the identity

$$\begin{aligned} \int d\alpha \langle \psi^l_\alpha, V^\mu_\alpha \psi^l_\alpha \rangle &= \sum_{1 \leq i < j \leq N} e_i e_j \int dX \int d\alpha \prod_{k=1}^N \chi^2_{u\alpha_k}(x_k/l) \psi^2(x_1, \dots, x_N) Y_\mu(x_i - x_j) \delta_{\alpha_i, \alpha_j} \\ &= \left\langle \psi, \sum_{1 \leq i < j \leq N} e_i e_j Y_\mu(x_i - x_j) h_l(x_i - x_j) \psi \right\rangle. \end{aligned} \tag{2.13}$$

Similarly, since for any $f \in C_0^\infty(\mathbb{R}^3)$

$$\langle f\chi, -\Delta(f\chi) \rangle = \langle f, -\Delta(\chi^2 f) \rangle + \langle f, |\nabla\chi|^2 f \rangle,$$

one has the following estimate for the kinetic energy with $C_0 = \sup_x |\nabla\chi|^2(x) < 3(\pi/2)^2$ (and recalling (1.9)):

$$\int d\alpha \langle \psi^l_\alpha, T\psi^l_\alpha \rangle \leq \langle \psi, T\psi \rangle + C_0 N l^{-2}. \tag{2.14}$$

We emphasize that, definition (2.12), particles in different boxes *do not interact*. Hence

$$\begin{aligned} &\left\langle \psi, \left[T + \sum_{1 \leq i < j \leq N} e_i e_j Y_\mu(x_i - x_j) h_l(x_i - x_j) \right] \psi \right\rangle + C_0 N l^{-2} \\ &\geq \int d\alpha \langle \psi^l_\alpha, (T + V^\mu_\alpha) \psi^l_\alpha \rangle \geq (\int d\alpha \|\psi^l_\alpha\|_2^2) \inf_{\sum n_\sigma = N} \sum_{\sigma \in \mathbb{Z}^3} E_{n_\sigma}^{\mu l} \\ &= \inf_{\sum n_\sigma = N} \sum_{\sigma \in \mathbb{Z}^3} E_{n_\sigma}^{\mu l}. \end{aligned} \tag{2.15}$$

Here $E_{n_\sigma}^{\mu l}$ is the ground state energy of a n_σ -particle system with Yukawa cutoff μ in a box of size l (see (2.2)). The sub-systems need to be neutral.

To complete the localization, one has to relate the potential $Y_\mu(z)h_l(z)$ to the Coulomb potential. Let

$$f_{\mu l}(z) \equiv a_0|z|^{-1} - Y_\mu(z)h_l(z). \tag{2.16}$$

The coefficient a_0 in front of r^{-1} is added for the purpose of normalization. Clearly, $f_{\mu l}(0) = a_0 \mu$. It will be shown in Lemma 2.1 below that $f_{\mu l}$ has a positive Fourier transform if $\mu l \geq C_3$ for some fixed constant C_3 . Hence by (2.15), (2.4) and (2.20)

$$E_N = a_0^{-2} E_{a_0, N} \geq a_0^{-2} \left[-C_0 N l^{-2} - \frac{1}{2} a_0 \mu N + \inf_{\sum n_\sigma = N} \sum_{\sigma \in \mathbb{Z}^3} E_{n_\sigma}^{\mu l} \right]. \quad (2.17)$$

Equation (2.17) is the localization estimate which we need to prove Theorem 1.2. Note that the correction terms are remarkably simple.

Let $l = N^{-\varepsilon}$ with ε some small number ($\varepsilon < 1/5$) and $\mu = C_3 N^\varepsilon$. Our goal is to apply Theorem 1.4 in each box.

Let ϕ be a n -particle wave function satisfying $\langle \phi, H_{2,n}^{\mu l} \phi \rangle \leq 0$. Then one has the trivial estimate (recall definition (2.2))

$$\frac{1}{2} \langle \phi, T \phi \rangle \leq -\frac{1}{2} \langle \phi, H_{2,n}^{\mu l} \phi \rangle \leq -\frac{1}{2} \inf \text{spec } H_{2,n}^{\mu l}. \quad (2.18)$$

But $H_{2,n}^{\mu l}$ can be bounded below by $-C_5 n^{5/3}$ [see (A.23) which is the stability of matter bound with Yukawa cutoff derived in the Appendix]. Hence the hypothesis of Theorem 1.4 is satisfied for each box with $l = N^{-\varepsilon}$ and $\gamma_\psi = N^{1/3-\varepsilon}$. We also have that $\sum_{\sigma \in \mathbb{Z}^3} (n_\sigma)^{7/5} \leq \left(\sum_{\sigma \in \mathbb{Z}^3} n_\sigma \right)^{7/5} = N^{7/5}$.

Let us now combine Theorem 1.4 with (2.17), temporarily ignoring the possibility that the particle number in some boxes may not be large. This yields

$$E_N \geq -0.30 a_0^{-2} N^{7/5} - a_0^{-2} C_0 N^{1+2\varepsilon} - C_3 a_0^{-1} N^{1+\varepsilon}. \quad (2.19)$$

To eliminate the last two terms as $N \rightarrow \infty$ we simply take $\varepsilon < 1/5$.

Despite the aforementioned problem about the particle number in each box, (2.19) is correct as $N \rightarrow \infty$. To prove this, note that in any box we can use $E_\sigma \geq -C_5 n_\sigma^{5/3}$. However, $\sup \{ \sum n_\sigma^{5/3} \mid \sum n_\sigma \leq N \text{ and } n_\sigma < N^p \} \leq N^{2p/3} \sum n_\sigma = N^{1+2p/3}$. If we take $0 < p < 3/5$, this shows that boxes with small particle number can be neglected.

Finally, to complete the proof of Theorem 1.2 we have to eliminate the a_0^{-2} factor in the first term in (2.19). This can be done as follows. Note that $a_0 = h(0)^3 = (\int g^4)^3$ with g given in (2.5), and g satisfies $\int g^2 = 1$. If g is replaced by $\eta(t) = 1$ for $|t| \leq \frac{1}{2}$ and $\eta(t) = 0$ otherwise, we would have $a_0 = 1$. But we cannot do this because $\int |\nabla g|^2$, the coefficient of the l^{-2} term in (2.14), would be infinite. Since $N l^{-2}$ is small on a scale of $N^{7/5}$, the remedy is to take $g \approx \eta$ and $\int |\nabla g|^2$ finite, but large. As $N \rightarrow \infty$, $g \rightarrow \eta$ and $a_0 \rightarrow 1$. Note that Lemma 2.1 does not depend on the special choice (2.5) we made for g .

This concludes the proof of Theorem 1.2 and we turn to Lemma 2.1.

Lemma 2.1. *Let $K: \mathbb{R}^3 \rightarrow \mathbb{R}$ be given by*

$$K(z) = r^{-1} \{ e^{-\nu r} - e^{-\omega r} h(z) \}$$

with $r = |z|$ and $\omega > \nu \geq 0$. Let h satisfy (i) h is a C^4 function of compact support; (ii) $h(z) = 1 + ar^2 + O(r^3)$ near $z = 0$. Let $h(z) = h(-z)$, so that K has a real Fourier transform. Then there is a constant C_3 (depending on h) such that if $\omega - \nu \geq C_3$ then

K has a positive Fourier transform and, moreover,

$$\sum_{1 \leq i < j \leq N} e_i e_j K(x_i - x_j) \geq \frac{1}{2}(v - \omega)N \tag{2.20}$$

for all $x_1, \dots, x_N \in \mathbb{R}^3$ and $e_i = \pm 1$.

Proof. Let $F(z) = [h(z) - 1 - ar^2]r^{-1}(1 + r^5)^{-1}$. $K(z)$ can thus be decomposed as

$$K(z) = Y_v(z) - Y_\omega(z) - are^{-\omega r} - (1 + r^5)e^{-\omega r}F(z).$$

The Fourier transforms of the first three terms are $4\pi/(p^2 + v^2)$, $-4\pi/(p^2 + \omega^2)$ and $-8\pi a(3\omega^2 - p^2)(p^2 + \omega^2)^{-3}$ respectively. For the last term note that $F(z)$ is of order r^2 and r^{-4} near the origin and near infinity, respectively and $\Delta^2 F(z)$ is of order r^{-2} and r^{-8} near the origin and near infinity, respectively. Therefore, ΔF and $\Delta^2 F \in L^1(\mathbb{R}^3)$ and hence (with $\hat{}$ denoting Fourier transform)

$$|(1 + p^2)^2 \hat{F}(p)| \leq 4\pi C_1$$

for some constant C_1 . But the Fourier transform of $(1 + r^5)e^{-\omega r}$ can be shown to satisfy $|((1 + r^5)e^{-\omega r})^\wedge| \leq 16\pi\omega(\omega^2 + p^2)^{-2}$ if $\omega \geq C_2$ for some constant C_2 . Hence

$$\begin{aligned} |[(1 + r^5)e^{-\omega r}F(z)]^\wedge| &= C_1 [8\pi\omega(\omega^2 + p^2)^{-2}] * [8\pi(1 + p^2)^{-2}] \\ &= C_1 [e^{-\omega r} \cdot e^{-r}]^\wedge = 8\pi C_1 (\omega + 1) [(\omega + 1)^2 + p^2]^{-2}. \end{aligned}$$

We can now put all these Fourier transform together to yield the estimate

$$\hat{K}(p) \geq 4\pi[(p^2 + v^2)^{-1} - (p^2 + \omega^2)^{-1} - 6|a|(p^2 + \omega^2)^{-2} - 2C_1(\omega + 1)(\omega^2 + p^2)^{-2}].$$

Hence $\hat{K}(p) \geq 0$ for all p if $\omega - v$ is large enough. To conclude the proof of Lemma 2.1, one only has to note the identity

$$\sum_{1 \leq i < j \leq N} e_i e_j K(x_i - x_j) = \frac{1}{2} \int \hat{K}(p) \left[\left| \sum_{j=1}^N e_j e^{ipx_j} \right|^2 - N \right] dp$$

which implies (2.20) since $\int \hat{K} = K(0) = \omega - v$. $\square$

Lemma 2.1 is applied to (2.16) with $v = 0$ and the requirement is that $\mu l \geq C_3$. However, our energy bound does not depend on the fact that we started with a Coulomb potential in (1.1). By the foregoing construction and Lemma 2.1 we have the following generalization of Theorem 1.2.

Theorem 2.1. *Let $e_i = \pm 1$ and let*

$$H_N^v = - \sum_{i=1}^N \Delta_i + \sum_{1 \leq i < j \leq N} e_i e_j Y_v(x_i - x_j) \tag{2.21}$$

be defined on $L^2(\mathbb{R}^{3N})$ with $Y_v(x) = |x|^{-1} \exp(-v|x|)$. v can depend on N , but suppose that as $N \rightarrow \infty$, $N^{-2/5}v \rightarrow 0$. Then, for sufficiently large N ,

$$H_N^v \geq -0.30N^{7/5}. \tag{2.22}$$

Proof. As in (2.16), we write $f_\mu = a_0 Y_v - Y_\mu h_l$. In order to apply our foregoing construction, the assumptions of Lemma 2.1 and Theorem 1.4 must be satisfied, namely $l(\mu - v) \geq C_3$, $l^{-1} \geq N^\epsilon$ and $l\mu N^{-1/5} < \infty$ as $N \rightarrow \infty$. On the other hand,

the correction terms resulting from the localization (cf. (2.17)) should be of lower order. Hence we must have $l^{-1} \leq o(N^{1/5})$ and $(\mu - \nu) \leq o(N^{2/5})$. It is easy to check that $\mu = \max(N^{1/5+\epsilon}, 2\nu)$ and $l = N^{1/5}\mu^{-1}$ satisfy all the requirements. $\square$

Returning to the Coulomb case, (1.1), we note the following *virial type theorem*.

Theorem 2.2 *Let the hypotheses be as in Theorem 1.2 and let ψ be any normalized (not necessarily symmetric) function in $L^2(\mathbf{R}^{3N})$. Let $K(\psi)$ and $P(\psi)$ denote the kinetic and potential energies of ψ (see (1.9) and $P(\psi) = \langle \psi, \sum e_i e_j |x_i - x_j|^{-1} \psi \rangle$). Then*

$$-P(\psi) \leq 2A^{1/2} N^{7/10} K(\psi)^{1/2}, \quad (2.23)$$

where $A = -N^{-7/5} \inf \text{spec}(H_N)$.

Proof. Replacing $\psi(x_i)$ by $\lambda^{3N/2} \psi(\lambda x_i)$ we find that $\lambda^2 K(\psi) + \lambda P(\psi) \geq -AN^{7/5}$. Then $-P(\psi) \geq \lambda^{-1} AN^{7/5} + \lambda K(\psi)$. Optimizing this with respect to λ yields (2.23). $\square$

III. The $\rho^{1/4}$ Law for Jellium

We shall prove Theorem 1.1 in this section by localizing the jellium Hamiltonian to a box of size $l = \rho_B^{-1/8}$. The localized Hamiltonian can thus be estimated by relating it to Theorem 2.1. In localizing the jellium Hamiltonian (1.2), one should be cautious about the fact that, after averaging over all translations, the coupling constant in the two-particle Coulomb interactions changes from 1 to a_0 [see (2.9)–(2.16)], while that of the particle-background remains unchanged. A straightforward localization as in Sect. II will fail to preserve the charge neutrality. We shall solve this difficulty by replacing the uniform background charge density, ρ_B , in each small box by a *non-uniform* background charge density which depends on the cutoff functions.

Let χ_Λ be the characteristic function of the big domain, Λ . For $\tau \in \mathbf{Z}^3$, $l > 0$, $\mu > 0$ and $\alpha = (\alpha_1, \dots, \alpha_N)$, and recalling (2.6), (2.7), *et seq.*, let

$$\rho_{B\tau}^{lu}(y) \equiv \chi_\Lambda(y) \chi_{u\tau}^2(y/l) \rho_B, \quad (3.1)$$

$$V_{B\tau\mu}^{lu}(x) \equiv \int Y_\mu(x-y) \rho_{B\tau}^{lu}(y) dy, \quad (3.2)$$

$$V_{B\mu}^l(x) \equiv \rho_B \int_\Lambda Y_\mu(x-y) h_l(x-y) dy. \quad (3.3)$$

Then using (2.11) one has the following definitions of $\mathcal{L}(\psi)$ and $\mathcal{R}(\psi)$ and localization estimate:

$$\begin{aligned} \mathcal{L}(\psi) &\equiv \int d\alpha \left\langle \psi_\alpha^l, \left[- \sum_{i=1}^N \left\{ \Delta_i + \sum_{\tau \in \mathbf{Z}^3} V_{B\tau\mu}^{lu}(x_i) \delta_{\alpha_i, \tau} \right\} \right. \right. \\ &\quad \left. \left. + \sum_{1 \leq i < j \leq N} Y_\mu(x_i - x_j) + \sum_{\tau \in \mathbf{Z}^3} \frac{1}{2} \int \rho_{B\tau}^{lu}(y) V_{B\tau\mu}^{lu}(y) dy \right] \psi_\alpha^l \right\rangle \\ &\leq \left\langle \psi, \left[- \sum_{i=1}^N \{ \Delta_i + V_{B\mu}^l(x_i) \} + \sum_{1 \leq i < j \leq N} Y_\mu(x_i - x_j) h_l(x_i - x_j) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \rho_B \int_\Lambda V_{B\mu}^l(y) dy \right] \psi \right\rangle + C_0 N l^{-2} \equiv \mathcal{R}(\psi). \end{aligned} \quad (3.4)$$

Equation (3.4) may appear to be complicated, but the proof is just a reordering of indices. Recall Eq. (2.6),

$$\begin{aligned} & \int d\alpha \left\langle \psi^l_\alpha, \sum_{j=1}^N \sum_{\tau \in \mathbb{Z}^3} V_{B\tau\mu}^{lu}(x_j) \delta_{\alpha_j, \tau} \psi^l_\alpha \right\rangle \\ &= \sum_{j=1}^N \left\langle \psi, \left[\rho_B \int_\Lambda dy \int d\alpha \sum_{\tau \in \mathbb{Z}^3} \delta_{\alpha_j, \tau} Y_\mu(x_j - y) \chi_{u\tau}^2(y/l) \prod_{i=1}^N \chi_{u\alpha_i}^2(x_i/l) \right] \psi \right\rangle \\ &= \sum_{j=1}^N \left\langle \psi, \rho_B \int_\Lambda dy Y_\mu(x_j - y) \left[\int_\Gamma du \sum_{\alpha_j} \chi_{u\alpha_j}^2(y/l) \chi_{u\alpha_j}^2(x_j/l) \right] \psi \right\rangle \\ &= \sum_{j=1}^N \langle \psi, V_{B\mu}^l(x_j) \psi \rangle, \\ & \int d\alpha \left\langle \psi^l_\alpha, \sum_{j=1}^N \sum_{\tau \in \mathbb{Z}^3} \iint \rho_{B\tau}^{lu}(y) \rho_{B\tau}^{lu}(y') Y_\mu(y - y') dy dy' \psi^l_\alpha \right\rangle \\ &= \left\langle \psi, \rho_B^2 \int_\Lambda dy \int_\Lambda dy' \left\{ \int_{\tau \in \mathbb{Z}^3} du \sum_{\alpha_j} \chi_{u\tau}^2(y/l) \chi_{u\tau}^2(y'/l) \right\} \sum_{\alpha_1, \dots, \alpha_N \in \mathbb{Z}^3} \prod_{j=1}^N \chi_{u\alpha_j}^2(x_j/l) \psi \right\rangle \\ &= \rho_B \int_\Lambda V_{B\mu}^l(y) dy. \end{aligned}$$

For the other terms in (3.4) one can use (2.13) and (2.14).

As in Sect. II, one can use the positive definiteness of $f_{\mu l}$ (Lemma 2.1) to yield the bound

$$\mathcal{R}(\psi) \leq \langle \psi, H_{N\Lambda}^J(a_0, \rho_B) \psi \rangle + C_0 N l^{-2} + N \mu a_0. \tag{3.5}$$

In (3.5) $H_{N\Lambda}^J(a_0, \rho_B)$ is the jellium Hamiltonian (1.2) but with all the potential energy terms multiplied by a_0 . To utilize (3.5) we have to relate the energy of $H_{N\Lambda}^J(a_0, \rho_B)$ to that of $H_{N\Lambda}^J$. By simple scaling this is given by

$$\inf \text{spec } H_{N, a_0 \Lambda}^J(1, \rho_B) = a_0^{-2} \inf \text{spec } H_{N\Lambda}^J(a_0, \rho_B a_0^3). \tag{3.6}$$

Let $l = \rho_B^{-1/8}$ and $\mu = C_6 \rho_B^{1/8}$. Then the last two terms in (3.5) are of order at most $N \rho_B^{1/4}$. To complete the proof of Theorem 1.1, one only has to show that $\mathcal{L}(\psi) \geq -C_7 N \rho_B^{1/4}$.

For each fixed τ and multi-index α , consider the localized Hamiltonian

$$-\sum_{j=1'}^N [\Delta_j \delta_{\tau\alpha_j} + V_{B\tau\mu}^{lu}(x_j) \delta_{\tau\alpha_j}] + \sum_{i < j} \delta_{\tau\alpha_i} \delta_{\tau\alpha_j} Y_\mu(x_i - x_j) + \tfrac{1}{2} \int \rho_{B\tau}^{lu}(y) V_{B\tau\mu}^{lu}(y) dy. \tag{3.7}$$

Our goal is to estimate the ground state of (3.7). Suppose $\alpha_1 = \alpha_2 = \dots = \alpha_M = \tau$ and $\alpha_j \neq \tau$ for $j > M$. Let $\rho_{B\tau}^{lu}(y) \equiv \tilde{\rho}_B(y)$ and $\tilde{V}_B^\mu \equiv Y_\mu * \tilde{\rho}_B$. Then (3.7) becomes

$$H_{BM}^\mu \equiv -\sum_{j=1}^M \{ \Delta_j + \tilde{V}_B^\mu(x_j) \} + \sum_{i < j}^M Y_\mu(x_i - x_j) + \tfrac{1}{2} \int \tilde{\rho}_B(y) \tilde{V}_B^\mu(y) dy. \tag{3.8}$$

Note that, by definition (3.1),

$$n_B \equiv \int \tilde{\rho}_B(y) dy \leq l^3 \rho_B. \tag{3.9}$$

Recall that the density function ρ_ϕ for an M -particle normalized wave function,

ϕ , is defined by

$$\rho_\phi(x) = M \int |\phi(x, x_2, \dots, x_M)|^2 dx_2 \cdots dx_M.$$

Therefore, if one defines

$$D_\mu(f) \equiv \frac{1}{2} \iint f(x) f(y) Y_\mu(x - y) dx dy, \quad (3.10)$$

and

$$\Omega(\phi) \equiv \left\langle \phi, \left[-\sum_{j=1}^M \Delta_j + \sum_{1 \leq i < j \leq M} Y_\mu(x_i - x_j) \right] \phi \right\rangle - D_\mu(\rho_\phi),$$

an easy calculation yields

$$\langle \phi, H_{BM}^\mu \phi \rangle = \Omega(\phi) + D_\mu(\rho_\phi - \tilde{\rho}_B). \quad (3.11)$$

Let $Q \equiv \int \rho_\phi - \int \tilde{\rho}_B$ be the value of the total charge in the small box. The following lemma is needed to bound the last term in (3.11).

Lemma 3.1 *Let $U = \{x \mid |x| \leq d\}$ be a ball of radius d and let $f: U \rightarrow \mathbf{R}$ be a (not necessarily positive) density satisfying $\int_U f = Q$. Then*

$$D_\mu(f) \geq \frac{1}{2} (Q^2/d) [1 + \mu d + \mu^2 d^2/3]^{-1}. \quad (3.12)$$

Proof. $D_\mu(f)$ can be written as

$$D_\mu(f) = \sup_h \int_U f h - \frac{1}{8\pi} \int_{\mathbf{R}^3} [|\nabla h|^2 + \mu^2 h^2].$$

To prove (3.10) we merely take (with $r = |x|$) $h(x) = \alpha$ for $r \leq d$ and $h(x) = \alpha d e^{-\mu(r-d)/r}$ for $r \geq d$. Then $\int f h = \alpha Q$. The $r < d$ part of the second integral is $\alpha^2 \mu^2 d^3/6$. The $r > d$ part can be calculated by integrating by parts, using $(-\Delta + \mu^2)h = 0$, and $d^2 h h'|_{r=d} = -\alpha^2 (\mu d^2 + d)$. This $r > d$ part is $\frac{1}{2} \alpha^2 d(1 + d\mu)$. Maximizing with respect to α yields (3.12). $\square$

Remark. Equation (3.12) is sharp when $\mu = 0$ or $\mu \rightarrow \infty$ with fixed d .

Returning to (3.11), recall that $l = \rho_B^{-1/8}$ and $\mu = C_6 \rho_B^{1/8}$. The $l \times l \times l$ cube fits into a ball of radius $d = 3^{1/2}l/2$. Applying (3.12) with $\mu d = \sqrt{3}C_6/2 \equiv C_7$ we find that, with $C_8 = 3^{-1/2}[1 + C_7 + C_7^2/3]^{-1}$,

$$D_\mu(\rho_\phi - \tilde{\rho}_B) \geq C_8 (M - n_B)^2 \rho_B^{1/8}. \quad (3.13)$$

Finally, we have to estimate $\Omega(\phi)$. For this purpose we introduce a “duplication of variables” trick. Consider the Hamiltonian on $L^2(\mathbf{R}^{3M})$.

$$H_{2M}^\mu = -\sum_{j=1}^{2M} \Delta_i + \sum_{1 \leq i < j \leq 2M} e_i e_j Y_\mu(x_i - x_j), \quad (3.14)$$

where $e_i = 1$ for $i \leq M$ and $e_i = -1$ for $i > M$. Let Φ be a normalized trial function defined by

$$\Phi(x_1, \dots, x_{2M}) = \phi(x_1, \dots, x_M) \phi(x_{M+1}, \dots, x_{2M}).$$

A simple calculation yields

$$\Omega(\phi) = \frac{1}{2} \langle \Phi, H_{2M}^\mu \Phi \rangle. \quad (3.15)$$

By (A.23) in the appendix,

$$\Omega(\phi) \geq -\frac{1}{2}(4.016)(2M)^{5/3}. \tag{3.16}$$

Let us divide the possible values of M into two cases.

(a) $M \leq \rho_B^{11/32}$. Here we use (3.11), (3.16) and $D_\mu(\rho_\phi - \bar{\rho}_B) \geq 0$ to conclude that

$$\langle \phi, H_{BM}^\mu \phi \rangle \geq -(6.375)M^{5/3} \geq -(6.375)M\rho_B^{11/48}. \tag{3.17}$$

(b) $M > \rho_B^{11/32}$. Here we use Theorem 2.1, (3.11), (3.13) and (3.15) to conclude that for large enough ρ_B

$$\langle \phi, H_{BM}^\mu \phi \rangle \geq -C_9(\rho_B)M^{7/5} + C_8(M - n_B)^2 \rho_B^{1/8}, \tag{3.18}$$

where $C_9(\rho_B) \rightarrow (0.30)2^{2/5}$ as $\rho_B \rightarrow \infty$. The statement “large enough ρ_B ” comes from the condition in Theorem 2.1 that $\mu M^{-2/5} \rightarrow 0$ as $M \rightarrow \infty$. By our assumption $\mu M^{-2/5} \leq C_6 \rho_B^{-1/80}$, and this goes to zero as $\rho_B \rightarrow \infty$. If, in (3.18), we recall that $n_B \leq \rho_B l^3 = \rho_B^{5/8}$ and $M > \rho_B^{11/32}$, it is easy to see that the right side of (3.18) is bounded below by $-C_{10}(\rho_B)M\rho_B^{1/4}$ and that $C_{10}(\rho_B) \rightarrow (0.30)2^{2/5}$ as $\rho_B \rightarrow \infty$.

Using these results (a) and (b), and summing over boxes, and recalling (3.6), we conclude that

$$E_0 \geq -[C_{11}(\rho_B)a_0^{-5/4} + C_0]N\rho_B^{1/4} \tag{3.19}$$

with $C_{11}(\rho_B) \rightarrow (0.30)2^{2/5}$ as $\rho_B \rightarrow \infty$. Recall from Sect. II that $a_0 = (3/4)^3$ and $C_0 < 3(\pi/2)^2$. Note that (3.19) holds for all N ; we did not take the limit $N \rightarrow \infty$ in deriving (3.19). With the bounds just given, the factor [] in (3.19) is 8.57 when $\rho_B \rightarrow \infty$.

This completes the proof of Theorem 1.1. This theorem can be generalized to the case of Yukawa potentials as in Theorem 2.1. It can also be generalized in another direction as follows.

Theorem 3.1. *Consider the modified jellium Hamiltonian with variable background charge density*

$$H_N^J = -\sum_{i=1}^N \{\Delta_i + V(x_i)\} + \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1} + \frac{1}{2} \int \rho(x)\rho(y)|x - y|^{-1} dx dy, \tag{3.20}$$

with $V(x) = \int \rho(y)|x - y|^{-1} dy$. The density ρ satisfies $-\infty < \rho(x) \leq \rho_B$ with $\rho_B \geq 0$. Then the ground state energy satisfies

$$E_0 \geq -A_9 N \rho_B^{1/4} \tag{3.21}$$

and A_9 satisfies the same bound as A_6 , given in Theorem 1.1.

The proof is an easy generalization of the one for constant $\rho(x) = \rho_B$ in Λ just given.

IV. Computation of Constants

Our main goal in this section is to obtain the constant 0.30 in the inequality (1.12). The calculation will consist of optimizing the methodology of [2]. We shall first make a heuristic calculation of the ground state energy E_0 of the Hamiltonian H_N

in (1.1) by using a modification of Bogoliubov's method, and will return to a proper proof of Theorem 1.4 later after Eq. (4.23). We find heuristically that $E_0 \sim -0.028N^{7/5}$ but we are not conjecturing that this constant is sharp. In the following we shall closely follow the notation of [2].

Heuristic Calculations. We consider the case $v = 2\pi/L$ (essentially the Coulomb case) and introduce periodic boundary conditions on Λ for the Hamiltonian H_N . Thus we have

$$\langle \psi, H_N \psi \rangle = \langle \psi, T \psi \rangle + \frac{1}{2} \sum_{k \in \mathbb{Z}^3} v(k) [\langle \psi, A_k^* A_k \psi \rangle - N]. \quad (4.1)$$

The operator T is the kinetic energy operator, which we write in the second quantized form (see [10] or [7], for example) as

$$T = L^{-2} 4\pi^2 \sum_{k \in \mathbb{Z}^3} k^2 [a_k^* a_k + b_k^* b_k]. \quad (4.2)$$

Here L is the length of a side of Λ . The operators a_k, b_k are annihilation operators corresponding to the two species of bosons and $k \in \mathbb{Z}^3$. The charge density operator A_k is given by

$$A_k = \sum_{n \in \mathbb{Z}^3} a_{n+k}^* a_n - b_{n+k}^* b_n. \quad (4.3)$$

The $v(k)$ is just the Fourier transform of the Yukawa potential Y_v (with $v = 2\pi/L$) divided by the volume of Λ . We take this value of $v > 0$ to avoid the singularity at $k = 0$. Thus $v(k)$ is given by

$$v(k) = [\pi L(|k|^2 + 1)]^{-1}. \quad (4.4)$$

In Bogoliubov's approximation one makes the ansatz

$$A_k \simeq \sum_{|m| \leq D\gamma} [S_{k,m}^* + T_{k,m}]. \quad (4.5)$$

Here, D and γ are constants which will be defined later in (4.14) and (4.23). The operators $S_{k,m}, T_{k,m}$ are defined as in (2.8) of [2] by

$$\begin{aligned} S_{k,m} &= \begin{cases} a_n^* a_{n+k} & \text{if } m = (n, 1) \\ -b_n^* b_{n+k} & \text{if } m = (n, -1) \end{cases}, \\ T_{k,m} &= \begin{cases} a_n^* a_{n-k} & \text{if } m = (n, 1) \\ -b_n^* b_{n-k} & \text{if } m = (n, -1) \end{cases}. \end{aligned} \quad (4.6)$$

In (4.6) $n \in \mathbb{Z}^3$ and ± 1 indicates the charge species; $|m|$ is defined to be $|n|$. The operators $a_n^\#, b_n^\#$ with $|n| \leq D\gamma$ are to be thought of as scalars subject to the constraint

$$\sum_{|n| \leq D\gamma} a_n^* a_n = \sum_{|n| \leq D\gamma} b_n^* b_n = \frac{N}{2}. \quad (4.7)$$

Hence if $|m| \leq D\gamma$ and $|k| > 2D\gamma$ the $S_{k,m}$ and $T_{k,m}$ are just annihilation operators. The expression (4.1) then becomes quadratic in creation and annihilation operators. One can compute its ground state energy exactly in the case when $D\gamma = 0$ but also

to a good approximation when $D\gamma > 0$. We do this by writing (4.1) as

$$\begin{aligned} \langle \psi, H_N^* \psi \rangle &= 4\pi^2 L^{-2} \sum_{|k| \leq D\gamma} k^2 \langle \psi, [a_k^* a_k + b_k^* b_k] \psi \rangle \\ &\quad + \frac{1}{2} \sum_{|k| > D\gamma N^\delta} v(k) I_k(\varepsilon_k, \psi) + \text{lower order terms}, \end{aligned} \tag{4.8}$$

where $\delta > 0$ is a small positive number. The expression $I_k(\varepsilon, \psi)$ is given, for general ε , by

$$\begin{aligned} I_k(\varepsilon, \psi) &= \left\langle \psi, \left[\varepsilon \sum_{|m| \leq D\gamma} [S_{k,m}^* S_{k,m} + T_{k,m}^* T_{k,m}] \right. \right. \\ &\quad \left. \left. + \left(\sum_{|m| \leq D\gamma} S_{k,m}^* + T_{k,m} \right)^* \left(\sum_{|m| \leq D\gamma} S_{k,m}^* + T_{k,m} \right) - N \right] \psi \right\rangle. \end{aligned} \tag{4.9}$$

The number ε_k is given by the formula

$$\varepsilon_k = 8\pi^2 k^2 [NL^2 v(k)]^{-1}. \tag{4.10}$$

One can compute exactly the ground state energy of $I_k(\varepsilon, \psi)$. It is given in [2] as

$$I_k(\varepsilon, \psi) \geq N \left\{ \left[\left(\frac{\varepsilon}{n_0} \right)^2 + 2 \left(\frac{\varepsilon}{n_0} \right) \right]^{1/2} - 1 - \frac{\varepsilon}{n_0} \right\}, \tag{4.11}$$

where

$$n_0 = \sum_{|m| \leq D\gamma} 1. \tag{4.12}$$

The right-hand side of (4.11) can be achieved if the numbers a_n, b_n satisfy

$$a_n^* a_n = b_n^* b_n = \frac{N}{n_0}, \quad |n| \leq D\gamma. \tag{4.13}$$

We shall take γ to be large, $\gamma \geq N^\delta$, and fix D to be the finite number

$$D = \pi^{-1} (5/12)^{1/2} = 0.645/\pi. \tag{4.14}$$

Then, to leading order of magnitude the ground state energy of the second sum in (4.8) is given by

$$\frac{-2N}{L} \left(\frac{NLn_0}{8\pi^3} \right)^{1/4} I, \tag{4.15}$$

where I is the integral

$$I = \int_0^\infty [1 + \xi^4 - (\xi^8 + 2\xi^4)^{1/2}] d\xi. \tag{4.16}$$

Observe that

$$0 < I < J = \int_0^\infty \frac{d\xi}{1 + 2\xi^4}. \tag{4.17}$$

Numerical values for I, J are given by

$$I = .806, \quad J = \pi 2^{1/4}/4 = 0.934. \tag{4.18}$$

The integral I can be expressed exactly in terms of elliptic integrals [7]. Since γ is large n_0 is given to leading order as

$$n_0 = 2 \int_{|x| \leq D\gamma} dx = \frac{8\pi}{3} (D\gamma)^3. \quad (4.19)$$

Hence (4.15) is given by

$$-B_0 \frac{N}{L} (NL\gamma^3)^{1/4}, \quad (4.20)$$

where

$$B_0 = 2(3\pi^2)^{-1/4} D^{3/4} I. \quad (4.21)$$

The formula (4.20) gives the second sum in (4.8) to leading order of magnitude. Next we need to calculate the first sum in (4.8) which is the macroscopic kinetic energy of the low lying occupied states subject to (4.13). This is clearly given to leading order of magnitude by

$$\begin{aligned} \frac{2N}{n_0} \sum_{|k| \leq D\gamma} \frac{4\pi^2}{L^2} |k|^2 &= \frac{2N}{n_0} \frac{4\pi^2}{L^2} \int_{|x| \leq D\gamma} |x|^2 dx \\ &= \left(\frac{12D^2\pi^2}{5} \right) \frac{N\gamma^2}{L^2} = \frac{N\gamma^2}{L^2}. \end{aligned} \quad (4.22)$$

The total energy of the system then, according to this calculation, is

$$\frac{12\pi^2}{5} N \left(\frac{D\gamma}{L} \right)^2 - 2(3\pi^2)^{-1/4} IN^{5/4} \left(\frac{D\gamma}{L} \right)^{3/4}. \quad (4.23)$$

If we minimize this expression with respect to $D\gamma/L$ we obtain the value $-0.028N^{7/5}$.

Proof of Theorem 1.4. In the following calculation we shall ignore all terms of lower order than $N^{7/5}$ since we are only concerned with proving the inequality (1.12) for $N \rightarrow \infty$. Let γ be the γ_ψ defined by (1.10). We can assume without loss of generality (by changing Λ to be a sufficiently large cube) that the δ in Theorem 1.4 is less than $\frac{1}{6}$ and that $\gamma \geq N^\delta$. Define μ by $\mu L = \max(N^{1/10}, \nu L)$ so that $\mu L \rightarrow \infty$ and $N^{-1/5} \mu L < \infty$ as $N \rightarrow \infty$. If $\nu L < N^{1/10}$, let us write $Y_\nu = (Y_\nu - Y_\mu) + Y_\mu$ and write

$$\langle \psi, H_N^\nu \psi \rangle = N^{-1/10} K(\psi) + P_\nu(\psi) - P_\mu(\psi) + (1 - N^{-1/10}) K(\psi) + P_\mu(\psi). \quad (4.24)$$

Here, $P_\nu(\psi)$ denotes the potential energy terms in H_N^ν with the Yukawa potential Y_ν . Since $Y_\nu - Y_\mu$ is positive definite, $P_\nu(\psi) - P_\mu(\psi) \geq -\frac{1}{2}(\mu - \nu)N \geq -\frac{1}{2}\mu N \geq -\frac{1}{2}N^{11/10}/L$. By the uncertainty principle in a box, $K(\psi) \geq CN/L^2$, whence the first three terms on the right side of (4.24) are at least $CN^{9/10}/L^2 - \frac{1}{2}N^{11/10}/L$. Minimizing this with respect to L , we find that these terms are bounded below by $-CN^{13/10} \gg -N^{7/5}$. For the last two terms on the right side of (4.24) we can clearly replace $N^{-1/10}$ by zero in the limit $N \rightarrow \infty$. Thus we need prove Theorem 1.4 only under the condition $\nu L \geq N^{1/10}$ and $N^{-1/5} \nu L$ bounded.

By taking Λ to be four times as big, we can suppose that ψ is supported in Q^N ,

where Q is a smaller cube of size $L/4$. We can then replace $Y_v(x)$ in H_N^v by its periodic extension

$$Y_v^L(x) = \sum_{n \in \mathbb{Z}^3} Y_v(x + nL), \tag{4.25}$$

because the difference in the two potential energies is at most $W(N) = N^2 L^{-1} e^{-vL/2}$ (with the factor N^2 coming from the number of interaction terms). Since $vL \geq N^{1/10}$, $W(N) < N^\varepsilon/L$ as $N \rightarrow \infty$ for every $\varepsilon > 0$. As before, we can borrow $N^{-1/10} K(\psi) \geq N^{9/10}/L^2$ to control $W(N)$.

Using Y_v^L in H_N^v , we then have that (4.1) is an identity provided that H_N^v is now understood to contain Y_v^L and provided (4.4) is replaced by

$$v(k) = [\pi L \{k^2 + (vL/2\pi)^2\}]^{-1}. \tag{4.26}$$

Clearly, $v(k) \geq [\pi L(|k|^2 + 1)]^{-1}$. Now we are ready to bound the various terms in (4.1).

First we bound the potential energy terms for $|k| < N^\delta D\gamma$ from below by

$$-\frac{1}{2} N \sum_{|k| \leq N^\delta D\gamma} v(k) \geq -CN^{1+\delta} D\gamma/L. \tag{4.27}$$

If, as before, we combine a small portion of the kinetic energy with (4.27) we obtain a lower bound which is lower order than $N^{7/5}$.

Next consider terms in the potential energy which have $|k| > N^\delta D\gamma$. We define $S_{k,m}$, and $T_{k,m}$ again as in (4.6) but this time for all m with $|m| \leq |k|/N^\delta$. Let us assume for the moment that the system is neutral so that the number of negative particles is $N/2$. We shall return to the nonneutral case after Eq. (4.68). Since $\gamma \leq N^{1/3-\delta}$, Lemma 2.2 of [2] becomes

$$\langle \psi, T\psi \rangle \geq \frac{4\pi^2}{NL^2} \sum_{|k| > N^\delta D\gamma} [1 - CN^{-\delta}] k^2 C_k(\psi), \tag{4.28}$$

and $C_k(\psi)$ is given by

$$C_k(\psi) = \sum_{r=0}^\infty 2^{-4r} \sum_{|m|}^r \langle \psi | S_{k,m}^* S_{k,m} + T_{k,m}^* T_{k,m} | \psi \rangle, \tag{4.29}$$

where $\sum_{|m|}^r$ is a sum over $(2^r - 1)N^{1/3-\delta/2} \leq |m| < (2^{r+1} - 1)N^{1/3-\delta/2}$. Note that the constant $4\pi^2/NL^2$ in (4.28) is better than that in [2]. This is due to the improved summation procedure in (4.29). Hence we have

$$\langle \psi, H_N^v \psi \rangle \geq \frac{1}{2} \sum_{|k| > N^\delta D\gamma} v(k) I_k(\varepsilon_k, \psi), \tag{4.30}$$

where

$$I_k(\varepsilon, \psi) = \varepsilon C_k(\psi) + \langle \psi | A_k^* A_k - N | \psi \rangle, \tag{4.31}$$

$$\varepsilon_k = 8\pi^2 k^2 [1 - CN^{-\delta}] [NL^2 v(k)]^{-1}. \tag{4.32}$$

Since the term $CN^{-\delta}$ in (4.32) is lower order, we shall ignore it in future computations.

Now let $a > 1$ be a positive number which we shall fix later. We define

$$\begin{aligned} n_0 &= \#\{m: |m| \leq D\gamma\}, \\ n_r &= \#\{m: a^{r-1}D\gamma < |m| \leq a^r D\gamma\}, \quad r = 1, 2, 3, \dots \end{aligned} \quad (4.33)$$

Evidently we have, to leading order,

$$\begin{aligned} n_0 &= \frac{8\pi}{3}(D\gamma)^3, \\ n_r &= \frac{8\pi}{3}(D\gamma)^3 a^{3(r-1)}[a^3 - 1]. \end{aligned} \quad (4.34)$$

We define N_r , $r = 1, 2, \dots$, to be the maximum possible number of particles, consistent with the given $K(\psi)$, such that $|k| > a^{r-1}D\gamma$. Thus we have

$$4\pi^2 N_r (a^{r-1}D\gamma)^2 L^{-2} = N\gamma^2 L^{-2}, \quad (4.35)$$

which yields

$$N_r = N[4\pi^2 (a^{r-1}D)^2]^{-1}. \quad (4.36)$$

We define $N_0 = N$.

The key inequality in [2] is

$$I_k(\varepsilon, \psi) \geq \sum_{r=0}^{\infty} [\alpha_r - (1 + \eta_r)N_r] + E_k. \quad (4.37)$$

The term E_k is a constant times the number of particles with momentum n satisfying $|n| \geq C|k|/N^\delta$. It follows that the expression

$$\sum_k v(k)E_k \quad (4.38)$$

can be combined with a small portion of the kinetic energy to yield a lower order term. We shall therefore concentrate on the sum on the right-hand side of (4.37). The η_r are defined as

$$\eta_r = \varepsilon/[n_r p_r], \quad (4.39)$$

and

$$p_r = 2^{4t} \quad \text{if} \quad (2^t - 1)N^{1/3-\delta} \leq a^r D\gamma < (2^{t+1} - 1)N^{1/3-\delta/2}. \quad (4.40)$$

The α_r are the positive roots of the polynomial equation (in μ)

$$1 + \sum_{r=0}^{\infty} N_r [\eta_r N_r - \mu]^{-1} + N_r [\eta_r N_r + \mu]^{-1} = 0. \quad (4.41)$$

We order the roots α_r in the following manner: Let α_0 be the unique root of (4.41) which has $\alpha_0 > \eta_0 N_0$. The roots α_r , $r = 1, 2, \dots$, are the unique roots of (4.41) which satisfy $\eta_{r-1} N_{r-1} > \alpha_r > \eta_r N_r$. We define $\beta_r(k)$ by

$$-\beta_r(k) = \alpha_r - (1 + \eta_r)N_r, \quad (4.42)$$

where α_r is determined from (4.41) after setting $\varepsilon = \varepsilon_k$ in the definition, (4.39), of η_r .

Let us define

$$\frac{1}{2} \sum_{|k| > N^\delta D\gamma} v(k) \beta_r(k) = B_r N (N L \gamma^3)^{1/4} / L. \quad (4.43)$$

Note that B_r is a constant plus correction terms which tend to zero as $NL\gamma^3 \rightarrow \infty$. In the following computation (cf. (4.66)), it will be found that $NL\gamma^3$ indeed tends to infinity, and thus we are able to neglect these correction terms.

If we define B by

$$B = \sum_{r=0}^{\infty} B_r, \tag{4.44}$$

we have that

$$\langle \psi, H_N^* \psi \rangle \geq -BN(NL\gamma^3)^{1/4}/L. \tag{4.45}$$

We need then to estimate β_r and B_r , for $r=0, 1, 2, \dots$. We first consider the case $r=0$. The root α_0 of (4.41) is clearly bounded below by the unique positive root of the equation

$$N_0[\eta_0 N_0 - \mu]^{-1} + N_0[\eta_0 N_0 + \mu]^{-1} + 1 = 0. \tag{4.46}$$

Hence we obtain

$$\beta_0 \leq N_0 \{1 + \eta_0 - [\eta_0^2 + 2\eta_0]^{1/2}\}. \tag{4.47}$$

Now substituting the values for $\beta_0(k)$ and performing the sum in (4.43) we obtain

$$B_0 = 2(3\pi^2)^{-1/4} D^{3/4} I. \tag{4.48}$$

In the calculation for (4.48) we have used the fact that $p_0 = 1$ in (4.39). In fact $p_r = 1$ provided $r \leq C \log N$, since $\gamma \geq N^{1/3-\delta}$. Note that (4.48) and (4.21) are identical.

Next, we wish to estimate β_r and B_r when $r = 1, 2, \dots$. Now α_r is bounded below by the unique root, μ , of the equation

$$1 + \sum_{j=0}^r N_j[\eta_j N_j - \mu]^{-1} + N_j[\eta_j N_j + \eta_r N_r]^{-1} = 0, \tag{4.49}$$

which lies in the interval $\eta_{r-1} N_{r-1} > \mu > \eta_r N_r$. Let $\alpha_{r,1}$ be the root of the polynomial equation which is the same as (4.49) except that the terms $N_j/(\eta_j N_j - \mu)$, $j=0, \dots, r-1$ are replaced by $N_j/(\eta_j N_j - \eta_r N_r)$, $j=0, \dots, r-1$. Thus $\alpha_{r,1}$ is *larger* than the corresponding root of (4.49). Next, let $\alpha_{r,2}$ be the root of the polynomial equation which is the same as (4.49) except that the terms $N_j/(\eta_j N_j - \mu)$, $j=0, \dots, r-1$ are replaced by $N_j/(\eta_j N_j - \alpha_{r,1})$. It is clear that $\alpha_{r,2}$ is *smaller* than the corresponding root of (4.49). We can define the quantities $\beta_{r,1}, B_{r,1}, B^1$ and $\beta_{r,2}, B_{r,2}, B^2$ to correspond to the roots $\alpha_{r,1}, \alpha_{r,2}$ respectively in exactly the same manner as β_r, B_r, B correspond to α_r .

We calculate $\alpha_{r,1}$. To do this we write the corresponding polynomial equation in the form

$$N_r[\eta_r N_r - \mu]^{-1} + 1 + h_{r,1}/(2\eta_r) = 0, \tag{4.50}$$

where $h_{r,1}$ is given by the equation

$$h_{r,1} = 1 + 2 \sum_{j=0}^{r-1} [\eta_j/\eta_r + N_r/N_j]^{-1} + [\eta_j/\eta_r - N_r/N_j]^{-1}. \tag{4.51}$$

From (4.50) it follows that

$$\beta_{r,1} = (1 + \eta_r)N_r - \alpha_{r,1} = N_r[1 + 2\eta_r/h_{r,1}]^{-1}. \tag{4.52}$$

We wish now to fix the values of a and D in an optimal way. We do this by making the approximation $h_{r,1} \simeq 1$ and optimizing the value of B^1 based on this. With this approximation we have, then, an approximate value for $B_{r,1}$ obtained by summing (4.52),

$$B_{r,1} \simeq \frac{1}{2} J [\pi (a^{r-1} D)]^{-2} \left[\frac{1}{3\pi^2} D^3 a^{3(r-1)} (a^3 - 1) \right]^{1/4}. \quad (4.53)$$

Summing (4.53) from $r = 1, \dots, \infty$ we have

$$\sum_{r=1}^{\infty} B_{r,1} \simeq \frac{1}{2} (3\pi^2)^{-1/4} \pi^{-2} J D^{-5/4} g(a), \quad (4.54)$$

where $g(a)$ is the function of a given by

$$g(a) = (a^3 - 1)^{1/4} a^{5/4} [a^{5/4} - 1]^{-1}. \quad (4.55)$$

We shall take $a = 2$, which is close to the minimum for g , and the corresponding value for g is $g(2) = 2.81$. From (4.48) and (4.54) we then have

$$B_0 + \sum_{r=1}^{\infty} B_{r,1} \simeq \frac{1}{2} (3\pi^2)^{-1/4} [4ID^{3/4} + 2.81\pi^{-2} J D^{-5/4}]. \quad (4.56)$$

The value of D is chosen to minimize the right side of (4.56). This yields the value

$$D = (1/\pi) [14.05J/12I]^{1/2} = 1.16/\pi. \quad (4.57)$$

It is of some interest to compare this value of D with the value of D given in (4.14), namely $D = .645/\pi$, which was used in the previous heuristic calculation. With D chosen as in (4.57), Eq. (4.56) yields

$$B_0 + \sum_{r=1}^{\infty} B_{r,1} \simeq 0.53. \quad (4.58)$$

Having fixed a and D we obtain an upper bound for B . The expression $h_{r,1}$ is given from (4.51) and (4.34), (4.36) as

$$\begin{aligned} h_{r,1} = 1 + 2 \sum_{j=1}^{r-1} \{ [a^{3(r-j)} + a^{2(j-r)}]^{-1} + [a^{3(r-j)} - a^{2(j-r)}]^{-1} \} \\ + 2 \left[a^{3(r-1)} (a^3 - 1) + \frac{a^{2(1-r)}}{4\pi^2 D^2} \right]^{-1} + 2 \left[a^{3(r-1)} (a^3 - 1) - \frac{a^{2(1-r)}}{4\pi^2 D^2} \right]^{-1}. \end{aligned} \quad (4.59)$$

It is easy to see from (4.59) that

$$1 < h_{r,1} < 5/3. \quad (4.60)$$

If we use the lower bound in (4.60) we obtain from (4.50) an upper bound on $\alpha_{r,1}$,

$$\alpha_{r,1} < \eta_r N_r + (1 + 1/2\eta_r)^{-1} N_r = \eta_r N_r [1 + 2(1 + 2\eta_r)^{-1}] \leq 3\eta_r N_r. \quad (4.61)$$

We may now use the upper bound (4.61) to obtain a *lower* bound on $\alpha_{r,2}$. In view of (4.61), $\alpha_{r,2}$ is bounded below by the root of the equation

$$N_r [\eta_r N_r - \mu]^{-1} + 1 + h_{r,2}/(2\eta_r) = 0, \quad (4.62)$$

where $h_{r,2}$ is given by the equation

$$h_{r,2} = 1 + 2 \sum_{j=0}^{r-1} [\eta_j/\eta_r + N_r/N_j]^{-1} + [\eta_j/\eta_r - 3N_r/N_j]^{-1}. \tag{4.63}$$

If we express $h_{r,2}$ in a similar fashion to (4.59) it is easy to conclude that $h_{r,2} < 5/3$. We conclude then that

$$\beta_{r,2} \leq N_r \left[1 + \frac{6\eta_r}{5} \right]^{-1}. \tag{4.64}$$

Hence from (4.58) we have

$$B \leq 0.53(5/3)^{1/4} = 0.60. \tag{4.65}$$

Thus (4.45) and (4.65) yield a lower bound on the energy, $\langle \psi, H_N^* \psi \rangle$, of the wave function, ψ , in terms of $\gamma = \gamma_\psi$ and L .

To obtain a lower bound on the energy in terms of N alone, we have to use the fact that γ, L and N are not really independent when the energy is negative and when the hypotheses of Theorem 1.4 are satisfied. To see this let us divide the kinetic energy into two parts. One part is estimated by using the definition of γ as $N\gamma^2/L^2$. The other part is put together with the potential energy and use is made of (4.45). In all, then, we have for any $\lambda, 0 < \lambda < 1$, the inequality

$$\langle \psi, H_N^* \psi \rangle \geq \frac{\lambda N \gamma^2}{L^2} - \frac{B}{(1-\lambda)^{1/4}} N^{5/4} \left(\frac{\gamma}{L} \right)^{3/4}. \tag{4.66}$$

The factor $(1-\lambda)^{-1/4}$ in (4.66) is obtained by applying scaling to (4.45). Minimizing (4.66) with respect to γ/L yields

$$\langle \psi, H_N^* \psi \rangle \geq -(5/8)(3/8)^{3/5} B^{8/5} N^{7/5} \lambda^{-3/5} (1-\lambda)^{-2/5}. \tag{4.67}$$

The maximum value of $h(\lambda) = \lambda^{3/5}(1-\lambda)^{2/5}$ for $0 < \lambda < 1$ is obtained at $\lambda = 3/5$ with $h(3/5) = .510$. Hence (4.67) yields

$$\langle \psi, H_N^* \psi \rangle \geq -0.30 N^{7/5}. \tag{4.68}$$

We have proven (4.68) under the assumption that H_N^* is the Hamiltonian of a *neutral* system. However for the argument of Sect. II to be valid we need to know that (4.68) holds even for nonneutral systems. The neutrality assumption entered in our calculations only in the inequality (4.28) and it did so in the following way. The estimate in Lemma 2.2 of [2] leads to a denominator $2 \max(N_+, N_-)$ instead of N in (4.28). It is only when $N_+ = N_- = N/2$ that we get (4.28). If the system is not neutral and the ratio of negative particles to the total number of particles is given by

$$\frac{N_-}{N} = \frac{(1-\xi)}{2}, \quad 0 \leq \xi \leq 1, \tag{4.69}$$

then the coefficient $4\pi^2/NL^2$ of the sum in (4.28) must be decreased to $4\pi^2/(1+\xi)NL^2$, which in turn leads to the inequality

$$\langle \psi, H_N^* \psi \rangle \geq -0.30(1+\xi)^{2/5} N^{7/5}. \tag{4.70}$$

The inequality (4.70) gives an $N^{7/5}$ lower bound for a nonneutral system which

has a slightly larger constant than the constant 0.30 for the neutral case. We wish to show that the constant 0.30 still holds for the nonneutral case in the situation where we apply this inequality in Sect. II. The Hamiltonian H_N^ν can be written as

$$H_N^\nu = W_N^\nu + H_N^\omega, \quad (4.71)$$

where W_N^ν is the N -body potential energy obtained from the function

$$r^{-1}(e^{-\nu r} - e^{-\omega r}) = \int_v^\omega e^{-ur} du. \quad (4.72)$$

We choose $\omega = \nu + N^{1/5}$. The inequality (4.70) applies to H_N^ω . In fact, the inequality becomes better since $\omega > \nu$, which implies that $\nu(k)$ becomes smaller. Our bound (4.70) is monotone in $\nu(k)$.

To bound W_N^ν from below, let us suppose the particles are fixed at points $x_1, \dots, x_N$ with the negative particles being at $x_i, i = 1, \dots, N_-$. We define a density $\rho(x)$, by

$$\rho(x) = \sum_{i=1}^N e_i \delta(x - x_i). \quad (4.73)$$

It is clear that

$$W_N^\nu = \frac{1}{2} \int_v^\omega du \int e^{-u|x-y|} \rho(x) \rho(y) dx dy - \frac{1}{2} N^{6/5}. \quad (4.74)$$

The following lemma and proof is due to Federbush [6, 3]. It can also be proved by the method of Lemma 3.1.

Lemma 4.1. *Let $\Lambda \subset \mathbf{R}^3$ be a cube of side length L . Let $f: \Lambda \rightarrow \mathbf{R}$ be a (not necessarily positive) density with $Q = \int_\Lambda f dx$. Let $\mu \geq 0$. Then there is a constant C_{14} independent of μ, f, L such that*

$$D_\mu(f) \equiv \frac{1}{2} \int f(x) f(y) \exp[-\mu|x-y|] dx dy \geq C_{14} Q^2 \mu L (1 + \mu^2 L^2)^{-2}.$$

Proof. Assume $f \in L^2(\Lambda)$ and write

$$\begin{aligned} D_\mu(f) &= \frac{1}{2} \langle f, e^{-\mu|x-y|} f \rangle = 4\pi\mu \langle f, (-\Delta + \mu^2)^{-2} f \rangle \\ &= 4\pi\mu \|(-\Delta + \mu^2)^{-1} f\|_2^2 \geq 4\pi\mu |\langle g, (-\Delta + \mu^2)^{-1} f \rangle|^2 / \|g\|_2^2 \\ &\geq 4\pi\mu |\langle h, f \rangle|^2 / \|(-\Delta + \mu^2)h\|_2^2, \end{aligned}$$

where g is any function in $L^2(\mathbf{R}^3)$ and where $g = (-\Delta + \mu^2)h$. Let H be a C^∞ function with $H(x) = 1$ for $|x| \leq 2$ and $H(x) = 0$ for $|x| \geq 3$. Finally, take $h(x) = H(x/L)$. $\square$

From (4.70) to (4.74) and Lemma 4.1 we conclude that

$$\langle \psi, H_N^\nu \psi \rangle \geq C_{14} \nu L (1 + \nu^2 L^2)^{-2} \xi^2 N^{11/5} - \frac{1}{2} N^{6/5} - 0.30 (1 + \xi)^{2/5} N^{7/5}. \quad (4.75)$$

The inequality (4.75) shows that (4.68) holds for large N , even in the nonneutral case, provide $\nu L N^{-1/5} < \infty$ as $N \rightarrow \infty$. (Recall that, as stated in the beginning, it is only necessary to prove Theorem 1.4 when $\nu L > N^{1/10}$.) This concludes the proof of Theorem 1.4.

V. The Nonneutral Case

Consider the Hamiltonian H_N of (1.1) and its generalization H_N^v of Theorem 2.1 or (2.21) acting on N_- negative particles and N_+ positive particles with $N_- \leq N_+, N_- + N_+ = N$. Our goal here is to generalize Theorems 1.2 and 2.1 as follows.

Theorem 5.1. *Let H_N^v be as in Theorem 2.1, and let there be N_- negative and N_+ positive particles with $N_- \leq N_+$. The parameter v can depend on N_- and N_+ but we suppose that $N_-^{-2/15} v \rightarrow 0$ as $N_- \rightarrow \infty$. (Note the difference from Theorem 2.1.) Then*

$$H_N^v \geq -A_5 N_-^{7/5} \tag{5.1}$$

for some constant, A_5 .

The proof follows the same lines as in Sect. II. One must modify it in two respects, however. First, it is necessary to prove that the interaction energy depends only on the number of negative particles. Second, we need to localize the kinetic energy in a somewhat different way than in Sect. II. Basically we only want to localize the kinetic energy of a positive particle if it lies in a box containing a negative particle. If we were to localize the kinetic energy of all positive particles, the cost in energy would be proportional to the number of positive particles and this of course could be much larger than $N_-^{7/5}$.

We solve the problem of the interaction energy in Lemma 5.3 below, but first we require the two preliminary Lemmas 5.1 and 5.2. The first is independently interesting.

Lemma 5.1. *Suppose that K and $L: \mathbf{R}^3 \rightarrow \mathbf{R}^+$ are two nonnegative functions (not necessarily symmetric) that satisfy the following (5.2), for some fixed, positive integer s ,*

$$sL(x) \geq K(y) \quad \text{whenever} \quad |x| \leq |y|. \tag{5.2}$$

Let $x_1, \dots, x_{N_-}$ and $y_1, \dots, y_{N_+}$ be points in $\mathbf{R}^3$ that satisfy

$$\sum_{i=1}^{N_-} K(y_j - x_i) - \sum_{1 \leq k \leq N_+, k \neq j} L(y_j - y_k) > 0 \tag{5.3}$$

for each $j = 1, \dots, N_+$. Then $N_+ \leq CsN_-$, where C is some universal geometric constant (60 will suffice).

Proof. We shall use the following geometric fact. There exists a finite set of closed, solid, circular cones in $\mathbf{R}^3$, each with apex at the origin and each with solid angle $\pi/3$ such that their union is all of $\mathbf{R}^3$. The minimum number of cones required for this is some integer C , and it is easy to see that $C \leq 60$. Denote these cones by $B_1, \dots, B_C$. Let Y denote the set of y_i points.

Now, without loss of generality, assume $x_1 = 0$. Let $Y_1 = \{y_i | y_i \in B_1\}$ be the points in B_1 , and let Z_1 be those s points in Y_1 which are closest to x_1 . (If there is an ambiguity, make an arbitrary choice; if Y_1 has fewer than s points then $Z_1 \equiv Y_1$.) Next, apply this process to the remainder $Y \setminus Y_1$ and thereby obtain Z_2 with respect to B_2 . Continuing in this way we obtain $Z_1, \dots, Z_C$ and $Y_1, \dots, Y_C$.

Let $Z = \bigcup_{i=1}^C Z_i$, whence Z has at most sC points.

Take $y_j \notin Z$ and consider the contribution to the left side of (5.3) coming from x_1 and Z . This contribution is

$$A_j = K(y_j) - \sum_{y_k \in Z} L(y_j - y_k).$$

We claim that $A_j \leq 0$. If $y_j \in B_\alpha$ then the second sum in A_j is not less than $\sum_{y_k \in Z_\alpha} L(y_j - y_k)$. But $|y_j - y_k|^2 = |y_j|^2 + |y_k|^2 - 2y_j \cdot y_k \leq |y_j|^2 + |y_k|^2 - |y_j||y_k| \leq |y_j|^2$, since $|y_k| \leq |y_j|$. Thus, $|y_j - y_k| \leq |y_j|$ and thus $sL(y_j - y_k) \geq K(y_j)$. Given that $y_j \in B_\alpha$, Z_α has s points and thus $A_j \leq 0$.

If we now remove x_1 and Z from the system we obtain a reduced system with a new $N_- = N_- - 1$ and with a new $N_+ \geq N_+ - sC$, and that satisfies (5.3) for all y_j in the new system. The construction can now be repeated with x_2 and then x_3 and so on until we obtain a final system with $N_- = 0$ and a final $\tilde{N}_+ \geq N_+ - sCN_-$. This clearly cannot satisfy (5.3) if $\tilde{N}_+ > 0$. $\square$

Lemma 5.2. Let $K: \mathbf{R}^3 \rightarrow \mathbf{R}^+$ be given as in Lemma 2.1 by

$$K(x) = r^{-1} \{e^{-vr} - e^{-\omega r} h(x)\}$$

with $r = |x|$ and $\omega > v \geq 0$. Here we assume only that $h: \mathbf{R}^3 \rightarrow \mathbf{R}$ satisfies (i) $-H \leq h(x) \leq 1$ for all x and some finite $H \geq 0$; (ii) h is continuous in some neighborhood of $x = 0$. Then there is a positive integer s such that

$$sK(x) \geq K(y) \quad \text{whenever} \quad |x| \leq |y|. \quad (5.4)$$

The integer s depends only on $\omega - v \equiv \rho$ and on h . For fixed h , s is a nonincreasing function of ρ .

Proof. For (5.4) we can restrict our attention to the case $v = 0, \omega = \rho$ because multiplication of this K by e^{-vr} only makes inequality (5.4) stronger. There is an $R > 0$ such that h is continuous in $B_R = \{x : |x| \leq R\}$. Since $K(x) \geq r^{-1} \{1 - e^{-\rho r}\}$, which is decreasing in r and since $K(x) \leq r^{-1} \{1 + He^{-\rho r}\}$, we have that $K(y)/K(x) \leq (1 + He^{-\rho r})/(1 - e^{-\rho r})$ with $r = |y|$. The maximum of this ratio for $r \geq R$ occurs at $r = R$ and is $s_1 = (1 + He^{-\rho R})/(1 - e^{-\rho R})$. Thus $s_1 K(y) \geq K(x)$ when $|x| \leq |y|$ and $|y| \geq R$. On the other hand, when $|x| \leq |y| \leq R$, consider the function $F_R(x, y) \equiv K(y)/K(x)$ defined in the closed set $T = \{(x, y) : |x| \leq |y| \leq R\}$. There are 2 cases. Case (i): $h(0) = 1$. Then K is continuous on B_R with $K(0) = \rho$. Moreover, $K(x) \geq r^{-1}(1 - e^{-\rho r})$ on B_R . Thus $F_R(x, y)$ is continuous and so has a maximum on T . Case (ii): $h(0) < 1$. Then $|x|K(x)$ is a continuous function on B_R and it is never zero, so $|x|K(x) \geq t$ for some $t > 0$. Hence, $F_R(x, y) \leq t^{-1}|x|K(y) \leq t^{-1}(1 + H)$. Thus, (5.4) is satisfied for any integer $s \geq \max\{s_1, \max_T F_R(x, y)\}$. To prove the monotonicity of s , consider $K(x)$ with $v \geq 0$ and $\omega = v + \rho$. Let $F(x, y) = K(y)/K(x)$. Since $s \geq 1$, we only consider x, y such that $F(x, y) \geq 1$. Hence $(\partial F / \partial \rho)(x, y) = [(\partial K / \partial \rho)(y) - F(\partial K / \partial \rho)(x)] / K(x) = -[|y|K(y) - e^{-v|y|} - |x|K(y) + Fe^{-v|x|}] / K(x)$. One concludes that $(\partial F / \partial \rho)(x, y) \leq 0$ if $|x| \leq |y|$ and if $F(x, y) \geq 1$. Hence s is monotone in ρ . $\square$

Lemma 5.3. Suppose $v, \rho, l \geq 0$ and let

$$f(x) = a_0 Y_v(x) - Y_{\rho+v}(x) h_l(x)$$

with $h_l(0) = a_0, h_l(x) = h(x/l)$ and with $h(x)$ given by (2.8). This f is a generalization of $f_{\mu l}$ given in (2.16). Given $x_1, \dots, x_N \in \mathbb{R}^3$ and $e_i = \pm 1$, let N_- (respectively N_+) be the number of e_i which are -1 (respectively $+1$). Assume that $N_- \leq N_+$. Finally, suppose that $\rho l \geq C_3$ (which is defined in Lemma 2.1 and which depends only on h), so that $\hat{f} \geq 0$. Then there is a constant C_{13} depending only on h and not on v, ρ, l , such that

$$\sum_{i < j} e_i e_j f(x_i - x_j) \geq -C_{13} \rho N_- . \tag{5.5}$$

Proof. Let W denote the left side of (5.5). Combining Lemmas 5.1 and 5.2, there is an s (which depends on h and on ρl (by scaling)) so that whenever $N_+ \geq CsN_-$ we can eliminate $N_+ - CsN_-$ positive particles without increasing W . Thus we can assume $N_+ \leq CsN_-$. Furthermore, this s can only decrease when ρl increases, so we can take s to be the value it has when $\rho l = C_3$ (which depends only on h). Thus s depends only on h . Now since $\hat{f} \geq 0$, we have $W \geq -\frac{1}{2}(N_+ + N_-)f(0) = -\frac{1}{2}(N_+ + N_-)\rho a_0 \geq -\frac{1}{2}(Cs + 1)\rho a_0 N_-$. $\square$

We return next to the problem of localizing the kinetic energy similarly to Sect. II. For any $\alpha = (u, \alpha_1, \dots, \alpha_N) \in \Gamma \times \mathbb{Z}^{3n}$ we define ψ_α^l as in (2.11). We adopt the convention that the negative particles are labelled $1, \dots, N_-$ and the positive particles are labelled $N_- + 1, \dots, N$. Let S_α be the α_i which correspond to the negative particles,

$$S_\alpha = \{\alpha_1, \dots, \alpha_{N_-}\} . \tag{5.6}$$

We denote by $\bar{S}_\alpha$ the set of nearest neighbors in $\mathbb{Z}^3$ of S_α , so

$$\bar{S}_\alpha = \{m \in \mathbb{Z}^3 \mid |m - \alpha_i| \leq \sqrt{3} \text{ for some } \alpha_i \in S_\alpha\} . \tag{5.7}$$

Let N_α be “the number of positive particles which lie in a box occupied by a negative particle” and $\bar{N}_\alpha$ “the number of positive particles which lie in the same box as a negative particle or a nearest neighbor of such box.” By this is meant

$$\begin{aligned} N_\alpha &= \#\{j > N_- \mid \alpha_j \in S_\alpha\}, \\ \bar{N}_\alpha &= \#\{j > N_- \mid \alpha_j \in \bar{S}_\alpha\}. \end{aligned} \tag{5.8}$$

The definition of $S_\alpha, \bar{S}_\alpha, N_\alpha, \bar{N}_\alpha$ depend only on $\alpha \in \mathbb{Z}^{3N}$. Finally we define the kinetic energy operator T_α (which also depends only on α) to be the kinetic energy of the negative particles plus the kinetic energy of “the positive particles which lie in a box occupied by a negative particle,” namely

$$T_\alpha = \sum_{i=1}^{N_-} -\Delta_i + \sum_{\{j > N_- \mid \alpha_j \in S_\alpha\}} -\Delta_j . \tag{5.9}$$

We then have the following lemma:

Lemma 5.4. *Let C_0 be the constant in (2.14). The kinetic energy is bounded below (recalling the definition of $\int d\alpha$ before (2.11)) as*

$$\langle \psi, T\psi \rangle \geq \frac{1}{2} \int d\alpha \langle \psi_\alpha^l, T_\alpha \psi_\alpha^l \rangle - C_0 l^{-2} [N_- + 27 \int \bar{N}_\alpha \|\psi_\alpha^l\|^2 d\alpha] . \tag{5.10}$$

Proof. We use (2.14) to bound $\langle \psi, -\Delta_i \psi \rangle$ below for $i \leq N_-$, namely

$$\left\langle \psi, \sum_{i=1}^{N_-} -\Delta_i \psi \right\rangle \geq \int d\alpha \left\langle \psi_\alpha^l, \sum_{i=1}^{N_-} -\Delta_i \psi_\alpha^l \right\rangle - C_0 l^{-2} N_- .$$

Now suppose $i > N_-$ and consider a fixed α . Then we have the inequality

$$\frac{1}{2} \int dx_i |\nabla_i (\chi_{u\alpha_i}(x_i/l) \psi)|^2 \leq \int dx_i |\nabla_i \chi_{u\alpha_i}(x_i/l)|^2 |\psi|^2 + \int dx_i \chi_{u\alpha_i}^2(x_i/l) |\nabla_i \psi|^2. \quad (5.11)$$

Now use the fact that

$$|\nabla_i \chi_{u\alpha_i}(x_i/l)|^2 \leq C_0 l^{-2} \sum_{\lambda \in \mathbb{Z}^3} g(\lambda - \alpha_i) \chi_{u\alpha_i}^2(x_i/l), \quad (5.12)$$

where $g(z)$ is the function $g(z) = 1$ if $|z| \leq \sqrt{3}$, $g(z) = 0$ if $|z| > \sqrt{3}$. Hence we have for all $i > N_-$,

$$\frac{1}{2} \langle \psi_\alpha^l, -\Delta_i \psi_\alpha^l \rangle \leq \int \prod_{j=1}^N \chi_{u\alpha_j}^2(x_j/l) |\nabla_i \psi|^2 dx + C_0 l^{-2} \sum_{\lambda \in \mathbb{Z}^3} g(\lambda - \alpha_i) \|\psi_{\alpha_1, \dots, \lambda, \dots, \alpha_N}^l\|^2, \quad (5.13)$$

with λ being in the i^{th} position in the last sum. For $i > N_-$ let T_α^i be the i^{th} term in the kinetic energy T_α in (5.9), namely $T_\alpha^i = -\Delta_i$ if $\alpha_i \in S_\alpha$, and $T_\alpha^i = 0$ otherwise. We have then from (5.13),

$$\frac{1}{2} \sum_{\alpha_i \in \mathbb{Z}^3} \langle \psi_\alpha^l, T_\alpha^i \psi_\alpha^l \rangle \leq \sum_{\alpha_i \in \mathbb{Z}^3} \int \prod_{j=1}^N \chi_{u\alpha_j}^2(x_j/l) |\nabla_i \psi|^2 dx + 27 C_0 l^{-2} \sum_{\alpha_i \in \bar{S}_\alpha} \|\psi_\alpha^l\|^2. \quad (5.14)$$

The number 27 is the number of nearest neighbors of a point in $\mathbb{Z}^3$. If we sum (5.14) with respect to all α_j for $j \neq i$, and then sum over i , and then integrate over $u \in \Gamma$, we obtain the inequality (5.10). $\square$

The following lemma is also needed for the proof of Theorem 5.1.

Lemma 5.5. *Let ψ_α^l be the localized wave function (2.11). Let V_α^μ be given by (2.12) and T_α by (5.9). Assume that $1 \leq \mu \leq N_-^{2/15}$. Then there is a constant $C = C(\mu l)$, depending only on μl such that, with the notation of (5.8), there is the estimate*

$$\frac{1}{2} \langle \psi_\alpha^l, T_\alpha \psi_\alpha^l \rangle + \langle \psi_\alpha^l, V_\alpha^\mu \psi_\alpha^l \rangle - 27 C_0 l^{-2} \bar{N}_\alpha \|\psi_\alpha^l\|^2 \geq -C(\mu l) N_-^{7/5} \|\psi_\alpha^l\|^2. \quad (5.15)$$

Proof. We analyze the left side of (5.15) similarly to (2.15). Since there is no interaction between boxes, the left-hand side of (5.15) is bounded below by

$$\left[\sum_{\sigma \in \mathbb{Z}^3} E_\sigma \right] \|\psi_\alpha^l\|^2, \quad (5.16)$$

where E_σ is the ground state energy of the following Hamiltonian, H_σ , depending on σ . There are three cases: $\sigma \in S_\alpha$, $\sigma \in \bar{S}_\alpha \setminus S_\alpha$ and $\sigma \notin \bar{S}_\alpha$. If $\sigma \in S_\alpha$ and n_σ of the i , $1 \leq i \leq N$, have $\alpha_i = \sigma$ with n_σ^- of these satisfying $i \leq N_-$, then H_σ is the Hamiltonian

$$H_\sigma = \frac{1}{2} T + V^\mu - 27 C_0 l^{-2} n_\sigma^+ \quad (5.17)$$

acting on n_σ particles in a box of size l , n_σ^- of which are negative, n_σ^+ positive. Here, $V^\mu = \sum_{i < j} e_i e_j Y_\mu(x_i - x_j)$. If $\sigma \in \bar{S}_\alpha \setminus S_\alpha$, then $n_\sigma^- = 0$ and H_σ is the Hamiltonian

$$H_\sigma = V^\mu - 27 C_0 l^{-2} n_\sigma \quad (5.18)$$

acting on n_σ positive particles in a box of size l . If $\sigma \notin \bar{S}_\alpha$ then H_σ is

$$H_\sigma = V^\mu \quad (5.19)$$

acting on n_σ positive particles.

We estimate the ground state energies E_σ . In the case of (5.19) we clearly have $E_\sigma \geq 0$. In the case of (5.17) we use (4.75). Taking into account the factor $\frac{1}{2}$ in (5.17), which gives a factor $2^{1/4}$ in the $N^{7/5}$ law (cf. (4.66)), and with $\tau \equiv \mu l$, (4.75) reads (with $v = \mu$ and $\omega = 2\mu$ instead of $\omega = \mu + N^{1/5}$)

$$H_\sigma = C_{14} \mu \tau (1 + \tau^2)^{-2} (n_\sigma^+ - n_\sigma^-)^2 - \frac{1}{2} \mu n_\sigma - (0.30) 2^{1/4} n_\sigma [\max(2n_\sigma^+, 2n_\sigma^-)]^{2/5} - 27 C_0 \mu^2 \tau^{-2} n_\sigma^+. \quad (5.20)$$

The quantities in (5.20) satisfy $n_\sigma^- \leq N_-$, τ is fixed, $1 \leq \mu \leq N_-^{2/15}$ and n_σ^+ is arbitrary. We minimize the left side of (5.20) with respect to n_σ^+ . One can show that

$$H_\sigma \geq -A(\tau) [\mu^2 n_\sigma^- + \mu^3 + (n_\sigma^-)^{7/5}] \quad (5.21)$$

for some A depending only on τ .

To bound (5.18), one simply notes that in this case

$$H_\sigma \geq \frac{1}{2} (\sqrt{3}l)^{-1} n_\sigma (n_\sigma - 1) \exp(-\sqrt{3}\tau) - 27 C_0 l^{-2} n_\sigma \geq -B(\tau) l^{-3}$$

for some $B(\tau)$ independent of n_σ . By using the fact that $1 \leq \mu \leq N_-^{2/15}$, one has $H_\sigma \geq -D(\tau) N_-^{2/5}$.

Now, putting together the bounds for (5.17), (5.18) and (5.19), we conclude that for some $F(\tau)$

$$\sum_{\sigma \in \mathbb{Z}^3} E_\sigma \geq -F(\tau) \sum_{\sigma}^{(1)} \{ \mu^2 n_\sigma^- + \mu^3 + n_\sigma^{-7/5} \} - F(\tau) \sum_{\sigma}^{(2)} N_-^{2/5},$$

where the first sum is over S_α and the second is over $\bar{S}_\alpha \setminus S_\alpha$. The number of points in S_α is N_- while the number of points in $\bar{S}_\alpha$ is at most $27 N_-$. Using the facts that $\mu \leq N_-^{2/15}$, $\sum_{\sigma}^{(1)} n_\sigma^- = N_-$, and the convexity of $n \rightarrow n^{7/5}$, the lemma is proved. $\square$

Proof of Theorem 5.1. Step 1. Starting with v , we define $\mu = N_-^{2/15}$, and $l = C_3 N_-^{2/15}$, where C_3 is given in Lemma 2.1. As in Sect. II we write $a_0 Y_v = f + Y_\mu h_l$, with $f = a_0 Y_v - Y_\mu h_l$ as in Lemma 5.3. By Lemma 5.3, the contribution to the potential energy from f is bounded below by $-C_{13}(\mu - v)N_- \geq -C_{13}N_-^{17/15}$ for large N . This can be neglected compared to $N_-^{7/5}$.

Step 2. Lemma 5.4 is used to localize the kinetic energy. The term $-C_0 l^{-2} N_-$ in (5.10) can be neglected since $l^{-2} = (C_3)^{-2} N_-^{4/15}$.

Step 3. The first and third terms on the right side of (5.10) is combined with the $Y_\mu h_l$ part of the potential energy. We localize this potential energy as in (2.13). The first and third terms of (5.10) plus the localized potential energy is just the left side of (5.15). To prove the theorem we merely have to sum the right side of (5.15) over α , but this is exactly $-C(C_3)N_-^{7/5}$ by the normalization condition on ψ . $\square$

Appendix: Thomas–Fermi Theory and the Stability of Matter with Yukawa Potentials

Our main goal here is to establish a lower bound to the energy and an upper bound to the kinetic energy for quantum mechanical particles interacting with

Yukawa, instead of Coulomb potentials. We consider N movable particles with charge -1 and coordinates $x_1, \dots, x_N \in \mathbf{R}^3$ and K fixed particles with coordinates $R_1, \dots, R_K \in \mathbf{R}^3$ and charges $z_1, \dots, z_K \geq 0$. The movable particles will be considered to be fermions with q spin states, so that $q = N$ corresponds to the boson case. The Hamiltonian is

$$H = - \sum_{i=1}^N \{ \Delta_i + V(x_i) \} + \sum_{1 \leq i < j \leq N} Y_\mu(x_i - x_j) + U, \quad (\text{A.1})$$

with

$$V(x) = \sum_{j=1}^K z_j Y_\mu(x - R_j),$$

$$U = \sum_{1 \leq i < j \leq K} z_i z_j Y_\mu(R_i - R_j). \quad (\text{A.2})$$

$Y_\mu(x) = |x|^{-1} \exp\{-\mu|x|\}$ is the Yukawa potential. It is positive definite and satisfies

$$(-\Delta + \mu^2) Y_\mu = 4\pi\delta. \quad (\text{A.3})$$

The energy is

$$E = \inf\{(\psi, H_n \psi) \mid \|\psi\|_2 = 1 \text{ and all } R_1, \dots, R_K\}. \quad (\text{A.4})$$

The method of [15] will be used, which means that we first have to examine the Thomas–Fermi (TF) functional

$$\mathcal{E}(\rho) = \frac{3}{5} q^{-2/3} \gamma \int \rho^{5/3}(x) dx - \int V(x) \rho(x) dx + \frac{1}{2} \iint \rho(x) \rho(y) Y_\mu(x - y) dx dy + U \quad (\text{A.5})$$

and corresponding energy

$$E^{\text{TF}} = \inf\{\mathcal{E}(\rho) \mid \rho \in L^{5/3} \cap L^1\}. \quad (\text{A.6})$$

Notice that in (A.6) we do not impose $\int \rho = N$. This constraint could easily be dealt with, but it is not needed in this paper.

One of our results will be that $E^{\text{TF}} - U$ is a monotone decreasing function of μ .

A. The Thomas–Fermi Problem. By the methods of [14], a minimizer exists for (A.6) and satisfies $\gamma q^{-2/3} \rho(x)^{2/3} = \max(\phi(x), 0)$ with

$$\phi(x) = V(x) - (Y_\mu * \rho)(x). \quad (\text{A.7})$$

Lemma A.1. $\phi(x) \geq 0$, all x , and therefore the TF equation becomes

$$\gamma q^{-2/3} \rho(x)^{2/3} = \phi(x). \quad (\text{A.8})$$

Proof. Let $B = \{x \mid \phi(x) < 0\}$. On B , $\rho(x) = 0$ and $R_i \notin B$, all i (because $\phi(R_i) = \infty$). Therefore $-\Delta\phi = -\mu^2\phi \geq 0$ on B , so ϕ is superharmonic on B . Since $\phi = 0$ on ∂B , $\phi \geq 0$ on B which implies that B is empty. $\square$

Lemma A.2. Let $z_1, \dots, z_K \geq 0$ and $\tilde{z}_1, z_2, \dots, z_K > 0$ be two sets of charges with $\tilde{z}_1 \geq z_1$. Then, for all x , $\tilde{\phi}(x) \geq \phi(x)$.

Proof. Let $\psi = \tilde{\phi} - \phi$ and $B = \{x \mid \psi < 0\}$. Clearly, $R_1 \notin B$. On B , $\tilde{\rho} \leq \rho$ so $(-\Delta + \mu^2)\psi = 4\pi(\rho - \tilde{\rho}) \geq 0$. Thus ψ is superharmonic on B and again B is empty. $\square$

Lemma A.3. Let $z_1, \dots, z_M > 0$ and $z_{M+1}, \dots, z_K > 0$ be two sets of charges located

at $R_1, \dots, R_K$. Then

$$E(z_1, \dots, z_K) > E(z_1, \dots, z_M) + E(z_{M+1}, \dots, z_K). \tag{A.9}$$

Proof. This is Teller’s theorem for the Yukawa potential and is proved as in [14] using Lemma A.2. $\square$

Lemma A.3 is given in [16, p. 237].

Next, we turn to the question of monotonicity with respect to μ .

Lemma A.4. Suppose $\mu_1 > \mu_2$, with given fixed charges $z_i > 0$ and locations R_i . Then $\phi_2(x) \geq \phi_1(x)$, for all x .

Proof. Let $\psi = \phi_2 - \phi_1$ and $B = \{x | \psi(x) < 0\}$. Then $(-\Delta + \mu_1^2)\phi_1(x) = \sum z_j \delta(x - R_j) - \rho_1(x)$. By subtracting these two equations, and using the fact that $\rho_1 > \rho_2$ on B , we find that $-\Delta\psi > \mu_1^2\phi_1 - \mu_2^2\phi_2 > 0$. Again, B is empty. $\square$

Let us define

$$N^c = \int \rho, \tag{A.10}$$

where ρ is the solution to (A.8). N^c is the maximum negative charge for the TF system (A.5).

Lemma A.5. If $\mu_1 > \mu_2$, with fixed z_i and R_i , then

$$N_1^c \leq N_2^c \quad \text{and} \quad E_1^{\text{TF}} - U_1 \geq E_2^{\text{TF}} - U_2. \tag{A.11}$$

Proof. $N_1^c \leq N_2^c$ is a trivial consequence of Lemma A.4 and (A.8). By multiplying (A.8) by ρ and integrating, we have that

$$E - U = -\frac{2}{5} \int V\rho - \frac{1}{10} \iint \rho(x)\rho(y) Y_\mu(x-y) dx dy. \tag{A.12}$$

Since $\mu_1 > \mu_2$, $\rho_1(x) \leq \rho_2(x)$ and $Y_{\mu_1}(x) < Y_{\mu_2}(x)$, for all x . This, together with (A.12), proves the lemma. $\square$

Let us now compare the Yukawa TF problem with the Coulomb TF problem, which corresponds to $\mu = 0$. For the Coulomb problem $N^c = Z \equiv \sum_{j=1}^K z_j$ [14]. By Lemmas A.3 and A.5 we have that

$$E^{\text{TF}} \geq \sum_{j=1}^K E^{\text{TF}, \text{atom}}(z_j) \geq \sum_{j=1}^K E_{\text{Coulomb}}^{\text{TF}, \text{atom}}(z_j). \tag{A.13}$$

The latter inequality follows from the fact that $U = 0$ for an atom. For the TF Coulomb atom [14], $E^{\text{TF}}(z) = -(3.679)\gamma^{-1}q^{2/3}z^{7/3}$. Thus, for the Yukawa problem,

$$E^{\text{TF}} \geq -(3.679)\gamma^{-1}q^{2/3} \sum_{j=1}^K z_j^{7/3}. \tag{A.14}$$

Another lower bound for $E^{\text{TF}, \text{atom}}(z)$ can be obtained by dropping the $\rho\rho Y_\mu$ term in (A.5). The resulting minimization problem is trivial: $q^{-2/3}\gamma\rho(x)^{2/3} = V(x) = zY_\mu(x)$ for an atom. Since $\int Y_\mu^{5/2} = 4\pi(2\pi/5\mu)^{1/2}$, (A.13) implies

$$E^{\text{TF}} \geq -4q\mu^{-1/2}\gamma^{-3/2}(2\pi/5)^{3/2} \sum_{j=1}^K z_j^{5/2}. \tag{A.15}$$

B. The Quantum-Mechanical Problem. Returning to the Hamiltonian in (A.1), we want to find a lower bound to $\langle \psi, H\psi \rangle$ for any normalized N -particle function, ψ . The one-particle density of ψ is defined by

$$\rho_\psi(x) = N \int |\psi(x, x_2, \dots, x_N)|^2 dx_2 \cdots dx_N, \quad (\text{A.16})$$

and $\langle \psi, H\psi \rangle$ will be bounded in terms of ρ_ψ .

To bound the particle-particle energy we use the trick in [15]. Consider (A.5) with $q = 1, K = N, \gamma = \delta$ (arbitrary), $R_i = x_i$ and $z_i = 1$ for $i = 1, \dots, N$. Then, inserting ρ_ψ in (A.1) and using (A.14),

$$\sum_{1 \leq i < j \leq N} Y_\mu(x_i - x_j) \geq \frac{1}{2} \iint \rho_\psi(x) \rho_\psi(y) Y_\mu(x - y) dx dy - \frac{3}{2} \delta \int \rho_\psi^{5/3} - 3.679N/\delta. \quad (\text{A.17})$$

To bound the kinetic energy, we use the bound in [15] (recall that $q = N$ for bosons):

$$K(\psi) = \left\langle \psi, - \sum_{i=1}^N \Delta_i \psi \right\rangle \geq K_3 N^{-2/3} \int \rho_\psi(x)^{5/3} dx. \quad (\text{A.18})$$

In [15], the constant K_3 is given as $\frac{3}{2}(3\pi/2)^{2/3} = 1.69$, but this constant was subsequently improved. The best bound at present is in [11] where it is shown that we can take $K_3 = 2.7709$.

Combining (A.17), (A.18) we have the following bound for any normalized ψ

$$\langle \psi, H\psi \rangle \geq \mathcal{E}(\rho_\psi) - (3.679)N/\delta, \quad (\text{A.19})$$

with $q = 1$ and $\gamma = \frac{5}{3}K_3 N^{-2/3} - \delta$ in (A.5). We choose

$$\delta = \frac{5}{3}K_3 N^{-1/6} \left[N^{1/2} + \left(\sum_{j=1}^K z_j^{7/3} \right)^{1/2} \right]^{-1}, \quad (\text{A.20})$$

which implies that $\gamma > 0$. Using the bound (A.14) we obtain

Theorem A.1. *With H given by (A.1), the following holds for all normalized ψ :*

$$\langle \psi, H\psi \rangle \geq -\frac{3}{5}(3.679)K_3^{-1}N^{2/3} \left[N^{1/2} + \left(\sum_{j=1}^K z_j^{7/3} \right)^{1/2} \right]^2 \quad (\text{A.21})$$

with $K_3 = 2.7709$.

The final task is to apply Theorem A.1 to H_N in (1.1). Suppose that K particles have $e_i = +1$ and M particles have $e_i = -1$ with $K + M = N$. By ignoring the positive kinetic energy of the positive particles, (A.21) can be used with $(N, K) \rightarrow (M, K)$. Alternatively, the roles of positive and negative particles can be interchanged, so we can also replace (N, K) in (A.21) by (K, M) . The two bounds can then be averaged and an expression of the form $\frac{1}{2}(K^{2/3} + M^{2/3})(K^{1/2} + M^{1/2})^2$ is obtained. However, given that $K + M = N$, $K^{2/3} + M^{2/3}$ has its maximum at $K = M = N/2$. So does $K^{1/2} + M^{1/2}$. Thus we have

Theorem A.2. *With H_N given by (1.1), the following holds for all normalized ψ .*

$$\langle \psi, H_N \psi \rangle \geq -1.004N^{5/3}. \quad (\text{A.22})$$

A virial type theorem, analogous to Theorem 2.2, can be obtained from (A.22). Another application is the following.

Theorem A.3. Suppose ψ is normalized and $\langle \psi, H_N \psi \rangle \leq 0$. Then

$$K(\psi) \leq 4.016N^{5/3}. \quad (\text{A.23})$$

Proof. $0 \geq K(\psi) + P(\psi) = \frac{1}{2}K(\psi) + \langle \psi, H_{N,1/2} \psi \rangle$ where $H_{N,1/2}$ is given by (1.1) but with Δ_i replaced by $\frac{1}{2}\Delta_i$. By scaling, the analogue of (A.22) is $\langle \psi, H_{N,1/2} \psi \rangle \geq -2(1.004)N^{5/3}$. $\square$

Acknowledgements. We are grateful to Michael Loss for many helpful discussions.

References

1. Bogoliubov, N. N.: On the theory of superfluidity. J. Phys. (USSR) **11**, 23–32 (1947)
2. Conlon, J. G.: The ground state energy of a Bose gas with Coulomb interaction II. Commun. Math. Phys. **108**, 363–374 (1987). See also part I, *ibid* **100**, 355–397 (1985)
3. Conlon, J. G.: The ground state energy of a classical gas. Commun. Math. Phys. **94**, 439–458 (1984)
4. Dyson, F. J.: Ground-state energy of a finite system of charged particles. J. Math. Phys. **8**, 1538–1545 (1967)
5. Dyson, F. J., Lenard, A.: Stability of matter I and II. J. Math. Phys. **8**, 423–434 (1967); *ibid* **9**, 698–711 (1968)
6. Federbush, P.: A new approach to the stability of matter problem II. J. Math. Phys. **16**, 706–709 (1975)
7. Foldy, L. L.: Charged boson gas. Phys. Rev. **124**, 649–651 (1961); Errata *ibid* **125**, 2208 (1962)
8. Girardeau, M.: Ground state of the charged Bose gas. Phys. Rev. **127**, 1809–1818 (1962)
9. Lieb, E. H.: The $N^{5/3}$ law for bosons. Phys. Lett. **70A**, 71–73 (1979)
10. Lieb, E. H.: The Bose fluid. In: Lectures in Theoretical Physics, Vol. VII C, pp. 175–224. Brittin, W. E. (ed.), Boulder: University of Colorado Press 1965
11. Lieb, E. H.: On characteristic exponents in turbulence. Commun. Math. Phys. **92**, 473–480 (1984)
12. Lieb, E. H., Narnhofer, H.: The thermodynamic limit for jellium. J. Stat. Phys. **12**, 291–310 (1975). Errata, *ibid* **14**, 465 (1976)
13. Lieb, E. H., Sakakura, A. Y.: Simplified approach to the ground-state energy of an imperfect Bose gas II. Charged Bose gas at high density. Phys. Rev. **133**, A899–A906 (1964)
14. Lieb, E. H., Simon, B.: The Thomas–Fermi theory of atoms, molecules and solids. Adv. Math. **23**, 22–116 (1977). See also, Lieb, E. H.: Thomas–Fermi and related theories of atoms and molecules. Rev. Mod. Phys. **53**, 603–641 (1981); Errata *ibid* **54**, 311 (1982)
15. Lieb, E. H., Thirring, W. E.: Bound for the kinetic energy of fermions which proves the stability of matter. Phys. Rev. Lett. **35**, 687–689 (1975). Errata, *ibid* **35**, 1116 (1975).
16. Thirring, W. E.: A course in mathematical physics, Vol. 4. Quantum mechanics of large systems. New York, Wien: Springer 1983

Communicated by A. Jaffe

Received October 21, 1987

Ground State Energy of the One-Component Charged Bose Gas*

Elliott H. Lieb^{1, **}, Jan Philip Solovej^{2, ***}

¹ Departments of Physics and Mathematics, Jadwin Hall, Princeton University, PO Box 708, Princeton, NJ 08544-0708, USA. E-mail: lieb@princeton.edu

² Department of Mathematics, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark. E-mail: solovej@math.ku.dk

Received: 23 August 2000 / Accepted: 5 October 2000

Dedicated to Leslie L. Foldy on the occasion of his 80th birthday

Abstract: The model considered here is the “jellium” model in which there is a uniform, fixed background with charge density $-\rho$ in a large volume V and in which $N = \rho V$ particles of electric charge $+e$ and mass m move – the whole system being neutral. In 1961 Foldy used Bogolubov’s 1947 method to investigate the ground state energy of this system for bosonic particles in the large ρ limit. He found that the energy per particle is $-0.402 r_s^{-3/4} m e^4 / \hbar^2$ in this limit, where $r_s = (3/4\pi\rho)^{1/3} e^2 m / \hbar^2$. Here we prove that this formula is correct, thereby validating, for the first time, at least one aspect of Bogolubov’s pairing theory of the Bose gas.

1. Introduction

Bogolubov’s 1947 pairing theory [B] for a Bose fluid was used by Foldy [F] in 1961 to calculate the ground state energy of the one-component plasma (also known as “jellium”) in the high density regime – which is the regime where the Bogolubov method was thought to be exact for this problem. Foldy’s result will be verified rigorously in this paper; to our knowledge, this is the first example of such a verification of Bogolubov’s theory in a three-dimensional system of bosonic particles.

Bogolubov proposed his approximate theory of the Bose fluid [B] in an attempt to explain the properties of liquid Helium. His main contribution was the concept of pairing of particles with momenta k and $-k$; these pairs are supposed to be the basic constituents of the ground state (apart from the macroscopic fraction of particles in the “condensate”, or $k = 0$ state) and they are the basic unit of the elementary excitations of the system. The pairing concept was later generalized to fermions, in which case the pairing was between

* © 2000 by the authors. This article may be reproduced in its entirety for non-commercial purposes.

** Work partially supported by U.S. National Science Foundation grant PHY98 20650-A01.

*** Work partially supported by EU TMR grant, by the Danish Research Foundation Center MaPhySto, and by a grant from the Danish Research Council.

particles having opposite momenta and, at the same time, opposite spin. Unfortunately, this appealing concept about the boson ground state has neither been verified rigorously in a 3-dimensional example, nor has it been conclusively verified experimentally (but pairing has been verified experimentally for superconducting electrons).

The simplest question that can be asked is the correctness of the prediction for the ground state energy (GSE). This, of course, can only be exact in a certain limit – the “weak coupling” limit. In the case of the charged Bose gas, interacting via Coulomb forces, this corresponds to the *high density* limit. In gases with short range forces the weak coupling limit corresponds to low density instead.

Our system has N bosonic particles with unit positive charge and coordinates x_j , and a uniformly negatively charged “background” in a large domain Ω of volume V . We are interested in the thermodynamic limit. A physical realization of this model is supposed to be a uniform electron sea in a solid, which forms the background, while the moveable “particles” are bosonic atomic nuclei. The particle number density is then $\rho = N/V$ and this number is also the charge density of the background, thus ensuring charge neutrality.

The Hamiltonian of the one-component plasma is

$$H = \frac{1}{2} \sum_{j=1}^N p_j^2 + U_{pp} + U_{pb} + U_{bb}, \quad (1)$$

where $p = -i\nabla$ is the momentum operator, $p^2 = -\Delta$, and the three potential energies, particle-particle, particle-background and background-background, are given by

$$U_{pp} = \sum_{1 \leq i < j \leq N} |x_i - x_j|^{-1}, \quad (2)$$

$$U_{pb} = -\rho \sum_{j=1}^N \int_{\Omega} |x_j - y|^{-1} d^3 y, \quad (3)$$

$$U_{bb} = \frac{1}{2} \rho^2 \int_{\Omega} \int_{\Omega} |x - y|^{-1} d^3 x d^3 y. \quad (4)$$

In our units $\hbar^2/m = 1$ and the charge is $e = 1$. The “natural” energy unit we use is two Rydbergs, $2Ry = me^4/\hbar^2$. It is customary to introduce the dimensionless quantity $r_s = (3/4\pi\rho)^{1/3} e^2 m/\hbar^2$. High density is small r_s .

The Coulomb potential is infinitely long-ranged and great care has to be taken because the finiteness of the energy per particle in the thermodynamic limit depends, ultimately, on delicate cancellations. The existence of the thermodynamic limit for a system of positive and negative particles, with the negative ones being fermions, was shown only in 1972 [LLe] (for the free energy, but the same proof works for the ground state energy). Oddly, the jellium case is technically a bit harder, and this was done in 1976 [LN] (for both bosons and fermions). One conclusion from this work is that neutrality (in the thermodynamic limit) will come about automatically – even if one does not assume it – provided one allows any excess charge to escape to infinity. In other words, given the background charge, the choice of a neutral number of particles has the lowest energy in the thermodynamic limit. A second point, as shown in [LN], is that e_0 is independent of the shape of the domain Ω provided the boundary is not too wild. For Coulomb systems this is not trivial and for real magnetic systems it is not even generally true. We take

advantage of this liberty and assume that our domain is a cube $[0, L] \times [0, L] \times [0, L]$ with $L^3 = V$.

We note the well-known fact that the lowest energy of H in (1) without any restriction about “statistics” (i.e., on the whole of $\otimes^N L^2(\mathbb{R}^3)$) is the same as for bosons, i.e., on the symmetric subspace of $\otimes^N L^2(\mathbb{R}^3)$. The fact that bosons have the lowest energy comes from the Perron–Frobenius Theorem applied to $-\Delta$.

Foldy’s calculation leads to the following theorem about the asymptotics of the energy for small r_s , which we call Foldy’s law.

Theorem 1.1 (Foldy’s Law). *Let E_0 denote the ground state energy, i.e., the bottom of the spectrum, of the Hamiltonian H acting in the Hilbert space $\otimes^N L^2(\mathbb{R}^3)$. We assume that $\Omega = [0, L] \times [0, L] \times [0, L]$. The ground state energy per particle, $e_0 = E_0/N$, in the thermodynamic limit $N, L \rightarrow \infty$ with $N/V = \rho$ fixed, in units of $me^4/\hbar^2$, is*

$$\begin{aligned} \lim_{V \rightarrow \infty} E_0/N &= e_0 = -0.40154 r_s^{-3/4} + o(\rho^{1/4}) \\ &= -0.40154 \left(\frac{4\pi}{3} \right)^{1/4} \rho^{1/4} + o(\rho^{1/4}), \end{aligned} \quad (5)$$

where the number -0.40154 is, in fact, the integral

$$A = \frac{1}{\pi} 6^{1/4} \int_0^\infty \left\{ p^2(p^4 + 2)^{1/2} - p^4 - 1 \right\} dp = -\frac{3^{1/4} \Gamma(3/4)}{5\sqrt{\pi} \Gamma(5/4)} \approx -0.40154. \quad (6)$$

Actually, our proof gives a result that is more general than Theorem 1.1. We allow the particle number N to be totally arbitrary, i.e., we do not require $N = \rho V$. Our lower bound is still given by (5), where now ρ refers to the background charge density.

In [F] 0.40154 is replaced by 0.80307 since the energy unit there is 1 Ry. The main result of our paper is to prove (5) by obtaining a lower bound on E_0 that agrees with the right side of (5). An upper bound to E_0 that agrees with (5) (to leading order) was given in 1962 by Girardeau [GM], using the variational method of himself and Arnowitt [GA]. Therefore, to verify (5) to leading order it is only necessary to construct a rigorous lower bound of this form and this will be done here. It has to be admitted, as explained below, that the problem that Foldy and Girardeau treat is slightly different from ours because of different boundary conditions and a concomitant different treatment of the background. We regard this difference as a technicality that should be cleared up one day, and do not hesitate to refer to the statement of 1.1 as a theorem.

Before giving our proof, let us remark on a few historical and conceptual points. Some of the early history about the Bose gas, can be found in the lecture notes [L].

Bogolubov’s analysis starts by assuming periodic boundary condition on the big box Ω and writing everything in momentum (i.e., Fourier) space. The values of the momentum, k are then discrete: $k = (2\pi/L)(m_1, m_2, m_3)$ with m_i an integer. A convenient tool for taking care of various $n!$ factors is to introduce second quantized operators $a_k^\#$ (where $a^\#$ denotes a or a^*), but it has to be understood that this is only a bookkeeping device. Almost all authors worked in momentum space, but this is neither necessary nor necessarily the most convenient representation (given that the calculations are not rigorous). Indeed, Foldy’s result was reproduced by a calculation entirely in x -space [LS]. Periodic boundary conditions are not physical, but that was always chosen for convenience in momentum space.

We shall instead let the particle move in the whole space, i.e., the operator H acts in the Hilbert space $L^2(\mathbb{R}^{3N})$, or rather, since we consider bosons, in the subspace

consisting of the N -fold fully symmetric tensor product of $L^2(\mathbb{R}^3)$. The background potential defined in (2) is however still localized in the cube Ω . We could also have confined the particles to Ω with Dirichlet boundary conditions. This would only raise the ground state energy and thus, for the lower bound, our setup is more general.

There is, however, a technical point that has to be considered when dealing with Coulomb forces. The background never appears in Foldy's calculation; he simply removes the $k = 0$ mode from the Fourier transform, v of the Coulomb potential (which is $v(k) = 4\pi|k|^{-2}$, but with k taking the discrete values mentioned above, so that we are thus dealing with a "periodized" Coulomb potential). The $k = 0$ elimination means that we set $v(0) = 0$, and this amounts to a subtraction of the average value of the potential – which is supposed to be a substitute for the effect of a neutralizing background. It does not seem to be a trivial matter to prove that this is equivalent to having a background, but it surely can be done. Since we do not wish to overload this paper, we leave this demonstration to another day. In any case the answers agree (in the sense that our rigorous lower bound agrees with Foldy's answer), as we prove here. If one accepts the idea that setting $v(0) = 0$ is equivalent to having a neutralizing background, then the ground state energy problem is finished because Girardeau shows [GM] that Foldy's result is a true upper bound within the context of the $v(0) = 0$ problem.

The potential energy is quartic in the operators $a_k^\#$. In Bogolubov's analysis only terms in which there are four or two $a_0^\#$ operators are retained. The operator a_0^* creates, and a_0 destroys particles with momentum 0 and such particles are the constituents of the "condensate". In general there are no terms with three $a_0^\#$ operators (by momentum conservation) and in Foldy's case there is also no four $a_0^\#$ term (because of the subtraction just mentioned).

For the usual short range potential there is a four $a_0^\#$ term and this is supposed to give the leading term in the energy, namely $e_0 = 4\pi\rho a$, where a is the "scattering length" of the two-body potential. Contrary to what would seem reasonable, this number, $4\pi\rho a$ is *not* the coefficient of the four $a_0^\#$ term, and to prove that $4\pi\rho a$ is, indeed, correct took some time. It was done in 1998 [LY] and the method employed in [LY] will play an essential role here. But it is important to be clear about the fact that the four $a_0^\#$, or "mean field" term is absent in the jellium case by virtue of charge neutrality. The leading term in this case presumably comes from the two $a_0^\#$ terms, and this is what we have to prove. For the short range case, on the other hand, it is already difficult enough to obtain the $4\pi\rho a$ energy that going beyond this to the two $a_0^\#$ terms is beyond the reach of rigorous analysis at the moment.

The Bogolubov ansatz presupposes the existence of Bose–Einstein condensation (BEC). That is, most of the particles are in the $k = 0$ mode and the few that are not come in pairs with momenta k and $-k$. Two things must be said about this. One is that the only case (known to us) in which one can verify the correctness of the Bogolubov picture at weak coupling is the *one-dimensional* delta-function gas [LLi] – in which case there is presumably *no* BEC (because of the low dimensionality). Nevertheless the Bogolubov picture remains correct at low density and the explanation of this seeming contradiction lies in the fact that BEC is not needed; what is really needed is a kind of condensation on a length scale that is long compared to relevant parameters, but which is fixed and need not be as large as the box length L . This was realized in [LY] and the main idea there was to decompose Ω into fixed-size boxes of appropriate length and use Neumann boundary conditions on these boxes (which can only lower the energy, and which is fine since we want a lower bound). We shall make a similar decomposition here, but, unlike the case

in [LY] where the potential is purely repulsive, we must deal here with the Coulomb potential and work hard to achieve the necessary cancellation.

The only case in which BEC has been proved to exist is in the hard core lattice gas at half-filling (equivalent to the spin-1/2 XY model) [KLS].

Weak coupling is sometimes said to be a “perturbation theory” regime, but this is not really so. In the one-dimensional case [LLi] the asymptotics near $\rho = 0$ is extremely difficult to deduce from the exact solution because the “perturbation” is singular. Nevertheless, the Bogolubov calculation gives it effortlessly, and this remains a mystery.

One way to get an excessively negative lower bound to e_0 for jellium is to ignore the kinetic energy. One can then show easily (by an argument due to Onsager) that the potential energy alone is bounded below by $e_0 \sim -\rho^{1/3}$. See [LN]. Thus, our goal is to show that the kinetic energy raises the energy to $-\rho^{1/4}$. This was done, in fact, in [CLY], but without achieving the correct coefficient $-0.803(4\pi/3)^{1/4}$. Oddly, the $-\rho^{1/4}$ law was proved in [CLY] by first showing that the *non-thermodynamic* $N^{7/5}$ law for a *two-component* bosonic plasma, as conjectured by Dyson [D], is correct.

The [CLY] paper contains an important innovation that will play a key role here. There, too, it was necessary to decompose $\mathbf{R}^3$ into boxes, but a way had to be found to eliminate the Coulomb interaction *between* different boxes. This was accomplished by *not* fixing the location of the boxes but rather averaging over all possible locations of the boxes. This “sliding localization” will play a key role here, too. This idea was expanded upon in [GG]. Thus, we shall have to consider only one finite box with the particles and the background charge in it independent of the rest of the system. However, a price will have to be paid for this luxury, namely it will not be entirely obvious that the number of particles we want to place in each box is the same for all boxes, i.e., $\rho\ell^3$, where ℓ is the length of box. Local neutrality, in other words, cannot be taken for granted. The analogous problem in [LY] is easier because no attractive potentials are present there. We solve this problem by choosing the number, n , in each box to be the number that gives the lowest energy in the box. This turns out to be close to $n = \rho\ell^3$, as we show and as we know from [LN] must be the case as $\ell \rightarrow \infty$.

Finally, let us remark on one bit of dimensional analysis that the reader should keep in mind. One should not conclude from (5) that a typical particle has energy $\rho^{1/4}$ and hence momentum $\rho^{1/8}$ or de Broglie wavelength $\rho^{-1/8}$. This is *not* the correct picture. Rather, a glance at the Bogolubov–Foldy calculation shows that the momenta of importance are of order $\rho^{-1/4}$, and the seeming paradox is resolved by noting that the number of excited particles (i.e., those not in the $k = 0$ condensate) is of order $N\rho^{-1/4}$. This means that we can, hopefully, localize particles to lengths as small as $\rho^{-1/4+\epsilon}$, and cut off the Coulomb potential at similar lengths, without damage, provided we do not disturb the condensate particles. It is this clear separation of scales that enables our asymptotic analysis to succeed.

2. Outline of the Proof

The proof of our Main Theorem 1.1 is rather complicated and somewhat hard to penetrate, so we present the following outline to guide the reader.

2.1. Section 3. Here we localize the system whose size is L into small boxes of size ℓ independent of L , but dependent on the intensive quantity ρ . Neumann boundary conditions for the Laplacian are used in order to ensure a lower bound to the energy. We

always think of operators in terms of quadratic forms and the Neumann Laplacian in a box Q is defined for all functions in $\psi \in L^2(Q)$ by the quadratic form

$$(\psi, -\Delta_{\text{Neumann}} \psi) = \int_Q |\nabla \psi(x)|^2 dx.$$

The lowest eigenfunction of the Neumann Laplacian is the constant function and this plays the role of the condensate state. This state not only minimizes the kinetic energy, but it is also consistent with neutralizing the background and thereby minimizing the Coulomb energy. The particles not in the condensate will be called “excited” particles.

To avoid localization errors we take $\ell \gg \rho^{-1/4}$, which is the relevant scale as we mentioned in the Introduction. The interaction among the boxes is controlled by using the sliding method of [CLY]. The result is that we have to consider only interactions among the particles and the background in each little box separately.

The N particles have to be distributed among the boxes in a way that minimizes the total energy. We can therefore not assume that each box is neutral. Instead of dealing with this distribution problem we do a simpler thing which is to choose the particle number in each little box so as to achieve the absolute minimum of the energy in that box. Since all boxes are equivalent this means that we take a common value n as the particle number in each box. The total particle number which is n times the number of boxes will not necessarily equal N , but this is of no consequence for a lower bound. We shall show later, however, that it equality is nearly achieved, i.e., the energy minimizing number n in each box is close to the value needed for neutrality.

2.2. Section 4. It will be important for us to replace the Coulomb potential by a cutoff Coulomb potential. There will be a short distance cutoff of the singularity at a distance r and a large distance cutoff of the tail at a distance R , with $r \leq R \ll \ell$. One of the unusual features of our proof is that r and R are not fixed once and for all, but are readjusted each time new information is gained about the error bounds.

In fact, already in Sect. 4 we give a simple preliminary bound on n by choosing $R \sim \rho^{-1/3}$, which is much smaller than the relevant scale $\rho^{-1/4}$, although the choice of R that we shall use at the end of the proof is of course much larger than $\rho^{-1/4}$, but less than ℓ .

2.3. Section 5. There are several terms in the Hamiltonian. There is the kinetic energy, which is non-zero only for the excited particles. The potential energy, which is a quartic term in the language of second quantization, has various terms according to the number of times the constant function appears. Since we do not have periodic boundary conditions we will not have the usual simplification caused by conservation of momentum, and the potential energy will be correspondingly more complicated than the usual expression found in textbooks.

In this section we give bounds on the different terms in the Hamiltonian and use these to get a first control on the condensation, i.e., a control on the number of particles $\hat{n}_+$ in each little box that are not in the condensate state.

The difficult point is that $\hat{n}_+$ is an operator that does not commute with the Hamiltonian and so it does not have a sharp value in the ground state. We give a simple preliminary bound on its average $\langle \hat{n}_+ \rangle$ in the ground state by again choosing $R \sim \rho^{-1/3}$. In order to control the condensation to an appropriate accuracy we shall eventually need not only a

bound on the average, $\langle \hat{n}_+ \rangle$, but also on the fluctuation, i.e., on $\langle \hat{n}_+^2 \rangle$. This will be done in Sect. 8 using a novel method developed in Appendix A for localizing off-diagonal matrices.

2.4. Section 6. The part of the potential energy that is most important is the part that is quadratic in the condensate operators $a_0^\#$ and quadratic in the excited variables $a_p^\#$ with $p \neq 0$. This, together with the kinetic energy, which is also quadratic in the $a_p^\#$, is the part of the Hamiltonian that leads to Foldy's law. Although we have not yet managed to eliminate the non-quadratic part up to this point we study the main "quadratic" part of the Hamiltonian. It is in this section that we essentially do Foldy's calculation.

It is not trivial to diagonalize the quadratic form and thereby reproduce Foldy's answer because there is no momentum conservation. In particular there is no simple relation between the resolvent of the Neumann Laplacian and the Coulomb kernel. The former is defined relative to the box and the latter is defined relative to the whole of $\mathbb{R}^3$. It is therefore necessary for us to localize the wavefunction in the little box away from the boundary. On such functions the boundary condition is of no importance and we can identify the kinetic energy with the Laplacian in all of $\mathbb{R}^3$. This allows us to have a simple relation between the Coulomb term and the kinetic energy term since the Coulomb kernel is in fact the resolvent of the Laplacian in all of $\mathbb{R}^3$.

When we cut off the wavefunction near the boundary we have to be very careful because we must not cut off the part corresponding to the particles in the condensate. To do so would give too large a localization energy. Rather, we cut off only functions with sufficiently large kinetic energy so that the localization energy is relatively small compared to the kinetic energy. The technical lemma needed for this is a double commutator inequality given in Appendix B.

2.5. Section 7. At this point we have bounds available for the quadratic part (from Sect. 6) and the annoying non-quadratic part (from Sect. 5) of the Hamiltonian. These depend on r , R , n , $\langle \hat{n}_+ \rangle$, and $\langle \hat{n}_+^2 \rangle$. We avail ourselves of the bounds previously obtained for n and $\langle \hat{n}_+ \rangle$ and now use our freedom to choose different values for r and R to bootstrap to the desired bounds on n and $\langle \hat{n}_+ \rangle$, i.e., we prove that there is almost neutrality and almost condensation in each little box.

2.6. Section 8. In order to control $\langle \hat{n}_+^2 \rangle$ we utilize, for the first time, the new method for localizing large matrices given in Appendix A. This method allows us to restrict to states with small fluctuations in $\hat{n}_+$, and thereby bound $\langle \hat{n}_+^2 \rangle$, provided we know that the terms that do not commute with $\hat{n}_+$ have sufficiently small expectation values. We then give bounds on these $\hat{n}_+$ "off-diagonal" terms. Unfortunately, these bounds are in terms of positive quantities coming from the Coulomb repulsion, but for which we actually do not have independent a-priori bounds. Normally, when proving a lower bound to a Hamiltonian, we can sometimes control error terms by absorbing them into positive terms in the Hamiltonian, which are then ignored. This may be done even when we do not have an a-priori bound on these positive terms. If we want to use Theorem A.1 in Appendix A, we will need an absolute bound on the "off-diagonal" terms and we can therefore not use the technique of absorbing them into the positive terms. The decision when to use the theorem in Appendix A or use the technique of absorption into positive terms is resolved in Sect. 9.

2.7. Section 9. Since we do not have an a-priori bound on the positive Coulomb terms as described above we are faced with a dichotomy. If the positive terms are, indeed, so large that enough terms can be controlled by them we do not need to use the localization technique of Appendix A to finish the proof of Foldy's law. The second possibility is that the positive terms are bounded in which case we can use this fact to control the terms that do commute with $\hat{n}_+$ and this allows us to use the localization technique in Appendix A to finish the proof of Foldy's law. Thus, the actual magnitude of the positive repulsion terms is unimportant for the derivation of Foldy's law.

3. Reduction to a Small Box

As described in the previous sections we shall localize the problem into smaller cubes of size $\ell \ll L$. We shall in fact choose ℓ as a function of ρ in such a way that $\rho^{1/4}\ell \rightarrow \infty$ as $\rho \rightarrow \infty$.

We shall localize the kinetic energy by using Neumann boundary conditions on the smaller boxes.

We shall first, however, describe how we may control the electrostatic interaction between the smaller boxes using the sliding technique of [CLY].

Let t , with $0 < t < 1/2$, be a parameter which we shall choose later to depend on ρ in such a way that $t \rightarrow 0$ as $\rho \rightarrow \infty$.

The choice of ℓ and t as functions of ρ will be made at the end of Sect. 9 when we complete the proof of Foldy's law.

Let $\chi \in C_0^\infty(\mathbb{R}^3)$ satisfy $\text{supp } \chi \subset [(-1+t)/2, (1-t)/2]^3$, $0 \leq \chi \leq 1$, $\chi(x) = 1$ for x in the smaller box $[(-1+2t)/2, (1-2t)/2]^3$, and $\chi(x) = \chi(-x)$. Assume that all m -th order derivatives of χ are bounded by $C_m t^{-m}$, where the constants C_m depend only on m and are, in particular, independent of t . Let $\chi_\ell(x) = \chi(x/\ell)$. Let $\eta = \sqrt{1-\chi}$. We shall assume that χ is defined such that η is also C^1 . Let $\eta_\ell(x) = \eta(x/\ell)$. Using χ we define the constant γ by $\gamma^{-1} = \int \chi(y)^2 dy$, and note that $1 \leq \gamma \leq (1-2t)^{-3}$. We also introduce the Yukawa potential $Y_\nu(x) = |x|^{-1}e^{-\nu|x|}$ for $\nu > 0$.

As a preliminary to the following Lemma 3.1 we quote Lemma 2.1 in [CLY].

Lemma. Let $K : \mathbb{R}^3 \rightarrow \mathbb{R}$ be given by

$$K(z) = r^{-1} \{e^{-\nu r} - e^{-\omega r} h(z)\}$$

with $r = |z|$ and $\omega > \nu \geq 0$. Let h satisfy (i) h is a C^4 function of compact support; (ii) $h(z) = 1 + ar^2 + O(r^3)$ near $z = 0$. Let $h(z) = h(-z)$, so that K has a real Fourier transform. Then there is a constant, C_3 (depending on h) such that if $\omega - \nu \geq C_3$ then K has a positive Fourier transform and, moreover,

$$\sum_{1 \leq i < j \leq N} e_i e_j K(x_i - x_j) \geq \frac{1}{2}(\nu - \omega)N$$

for all $x_1, \dots, x_N \in \mathbb{R}^3$ and all $e_i = \pm 1$.

Lemma 3.1 (Electrostatic decoupling of boxes using sliding). *There exists a function of the form $\omega(t) = Ct^{-4}$ (we assume that $\omega(t) \geq 1$ for $t < 1/2$) and a constant γ with $1 \leq \gamma \leq (1-2t)^{-3}$ such that if we set*

$$w(x, y) = \chi_\ell(x) Y_{\omega(t)/\ell}(x - y) \chi_\ell(y) \quad (7)$$

then the potential energy satisfies

$$\begin{aligned}
 & U_{pp} + U_{pb} + U_{bb} \\
 & \geq \gamma \sum_{\lambda \in \mathbb{Z}^3} \int_{\mu \in [-\frac{1}{2}, \frac{1}{2}]^3} d\mu \left\{ \sum_{1 \leq i < j \leq N} w(x_i + (\mu + \lambda)\ell, x_j + (\mu + \lambda)\ell) \right. \\
 & \quad - \rho \sum_{j=1}^N \int_{\Omega} w(x_j + (\mu + \lambda)\ell, y + (\mu + \lambda)\ell) dy \\
 & \quad \left. + \frac{1}{2} \rho^2 \iint_{\Omega \times \Omega} w(x + (\mu + \lambda)\ell, y + (\mu + \lambda)\ell) dx dy \right\} - \frac{\omega(t)N}{2\ell}.
 \end{aligned}$$

Proof. We calculate

$$\begin{aligned}
 & \sum_{\lambda \in \mathbb{Z}^3} \int_{\mu \in [-1/2, 1/2]^3} d\mu \gamma \chi(x + (\mu + \lambda)) Y_{\omega}(x - y) \chi(y + (\mu + \lambda)) \\
 & = \int \gamma \chi(x + z) Y_{\omega}(x - y) \chi(y + z) dz = h(x - y) Y_{\omega}(x - y),
 \end{aligned}$$

where we have set $h = \gamma \chi * \chi$. Note that $h(0) = 1$ and that h satisfies all the assumptions in Lemma 2.1 in [CLY]. We then conclude from Lemma 2.1 in [CLY] that the Fourier transform of the function $F(x) = |x|^{-1} - h(x) Y_{\omega(t)}(x)$ is non-negative, where ω is a function such that $\omega(t) \rightarrow \infty$ as $t \rightarrow 0$. [The detailed bounds from [CLY] show that we may in fact choose $\omega(t) = Ct^{-4}$, since $\omega(t)$ has to control the 4th derivative of h .] Note, moreover, that $\lim_{x \rightarrow 0} F(x) = \omega(t)$. Hence

$$\begin{aligned}
 & \sum_{1 \leq i < j \leq N} F(y_i - y_j) - \rho \sum_{j=1}^N \int_{\ell^{-1}\Omega} F(y_j - y) dy \\
 & \quad + \frac{1}{2} \rho^2 \iint_{\ell^{-1}\Omega \times \ell^{-1}\Omega} F(x - y) dx dy \geq -\frac{N\omega(t)}{2}.
 \end{aligned}$$

The lemma follows by writing $|x|^{-1} = F(x) + h(y) Y_{\omega(t)}(x)$ and by rescaling from boxes of size 1 to boxes of size ℓ . $\square$

As explained above we shall choose the parameters t and ℓ as functions of ρ at the very end of the proof. We shall choose them in such a way that $t \rightarrow 0$ and $\rho^{1/4}\ell \rightarrow \infty$ as $\rho \rightarrow \infty$. Moreover, we will have conditions of the form

$$\rho^{-\tau}(\rho^{1/4}\ell) \rightarrow 0, \quad \text{and } t^v(\rho^{1/4}\ell) \rightarrow \infty$$

as $\rho \rightarrow \infty$, where τ, v are universal constants.

Consider now the n -particle Hamiltonian

$$H_{\mu, \lambda}^n = -\frac{1}{2} \sum_{j=1}^n \Delta_{Q_{\mu, \lambda}}^{(j)} + \gamma W_{\mu, \lambda}, \quad (8)$$

where we have introduced the Neumann Laplacian $\Delta_{Q_{\mu,\lambda}}^{(j)}$ of the cube $Q_{\mu,\lambda} = (\mu + \lambda)\ell + [-\frac{1}{2}\ell, \frac{1}{2}\ell]^3$ and the potential

$$\begin{aligned} W_{\mu,\lambda}(x_1, \dots, x_n) = & \sum_{1 \leq i < j \leq n} w(x_i + (\mu + \lambda)\ell, x_j + (\mu + \lambda)\ell) \\ & - \rho \sum_{j=1}^n \int_{\Omega} w(x_j + (\mu + \lambda)\ell, y + (\mu + \lambda)\ell) dy \\ & + \frac{1}{2}\rho^2 \iint_{\Omega \times \Omega} w(x + (\mu + \lambda)\ell, y + (\mu + \lambda)\ell) dx dy. \end{aligned}$$

Lemma 3.2 (Decoupling of boxes). *Let $E_{\mu,\lambda}^n$ be the ground state energy of the Hamiltonian $H_{\mu,\lambda}^n$ given in (8) considered as a bosonic Hamiltonian. The ground state energy E_0 of the Hamiltonian H in (1) is then bounded below as*

$$E_0 \geq \sum_{\lambda \in \mathbb{Z}^3} \int_{\mu \in [-\frac{1}{2}, \frac{1}{2}]^3} \inf_{1 \leq n \leq N} E_{\mu,\lambda}^n d\mu - \frac{\omega(t)N}{2\ell}.$$

Proof. If $\Psi(x_1, \dots, x_N) \in L^2(\mathbb{R}^{3N})$ is a symmetric function. Then

$$(\Psi, H\Psi) \geq \sum_{\lambda \in \mathbb{Z}^3} \int_{\mu \in [-\frac{1}{2}, \frac{1}{2}]^3} (\Psi, \tilde{H}_{\mu,\lambda}\Psi) d\mu - \frac{\omega(t)N}{2\ell},$$

where

$$\begin{aligned} (\Psi, \tilde{H}_{\mu,\lambda}\Psi) = & \sum_{j=1}^N \int_{x_j \in Q_{\mu,\lambda}} |\nabla_j \Psi(x_1, \dots, x_N)|^2 dx_1 \dots dx_N \\ & + \gamma \int W_{\mu,\lambda}(x_1, \dots, x_N) |\Psi(x_1, \dots, x_N)|^2 dx_1 \dots dx_N. \end{aligned}$$

The lemma follows since it is clear that $(\Psi, \tilde{H}_{\mu,\lambda}\Psi) \geq \inf_{1 \leq n \leq N} E_{\mu,\lambda}^n$. $\square$

For given μ the Hamiltonians $H_{\mu,\lambda}^n$ fall in three groups depending on λ . The first kind for which $Q_{\lambda,\mu} \cap \Omega = \emptyset$. They describe boxes with no background. The optimal energy for these boxes are clearly achieved for $n = 0$. The second kind for which $Q_{\lambda,\mu} \subset \Omega$. These Hamiltonians are all unitarily equivalent to γH_{ℓ}^n , where

$$\begin{aligned} H_{\ell}^n = & \sum_{j=1}^n \left(-\frac{1}{2}\gamma^{-1} \Delta_{\ell,j} - \rho \int w(x_j, y) dy \right) \\ & + \sum_{1 \leq i < j \leq n} w(x_i, x_j) + \frac{1}{2}\rho^2 \iint w(x, y) dx dy, \end{aligned} \tag{9}$$

where $-\Delta_{\ell}$ is the Neumann Laplacian for the cube $[-\ell/2, \ell/2]^3$. Finally, there are operators of the third kind for which $Q_{\mu,\lambda}$ intersects both Ω and its complement. In

this case the particles only see part of the background. If we artificially add the missing background only the last term in the potential $W_{\mu,\lambda}$ increases. (The first term does not change and the second can only decrease.) In fact it will increase by no more than

$$\frac{1}{2}\rho^2 \iint w(x, y) dx dy \leq \frac{1}{2}\rho^2 \iint_{\substack{x \in [-\ell/2, \ell/2]^3 \\ y \in [-\ell/2, \ell/2]^3}} |x - y|^{-1} dx dy \leq C\rho^2 \ell^5.$$

Thus the operator $H_{\mu,\lambda}^n$ of the third kind are bounded below by an operator which is unitarily equivalent to $\gamma H_\ell^n - C\rho^2 \ell^5$.

We now note that the number of boxes of the third kind is bounded above by $C(L/\ell)^2$. The total number of boxes of the second or third kind is bounded above by $(L + \ell)^3/\ell^3 = (1 + L/\ell)^3$.

We have therefore proved the following result.

Lemma 3.3 (Reduction to one small box). *The ground state energy E_0 of the Hamiltonian H in (1) is bounded below as*

$$E_0 \geq (1 + L/\ell)^3 \gamma \inf_{1 \leq n \leq N} \inf \text{Spec } H_\ell^n - C(L/\ell)^2 \rho^2 \ell^5 - \frac{\omega(t)N}{2\ell},$$

where H_ℓ^n is the Hamiltonian defined in (9).

In the rest of the paper we shall study the Hamiltonian (9).

4. Long and Short Distance Cutoffs in the Potential

The potential in the Hamiltonian (9) is w given in (7). Our aim in this section to replace w by a function that has long and short distance cutoffs.

We shall replace the function w by

$$w_{r,R}(x, y) = \chi_\ell(x) V_{r,R}(x - y) \chi_\ell(y), \quad (10)$$

where

$$V_{r,R}(x) = Y_{R^{-1}}(x) - Y_{r^{-1}}(x) = \frac{e^{-|x|/R} - e^{-|x|/r}}{|x|}. \quad (11)$$

Here $0 < r \leq R \leq \omega(t)^{-1}\ell$. Note that for $x \ll r$ then $V_{r,R}(x) \approx r^{-1} - R^{-1}$ and for $|x| \gg R$ then $V_{r,R}(x) \approx |x|^{-1} e^{-|x|/R}$.

In this section we shall bound the effect of replacing w by $w_{r,R}$. We shall not fix the cutoffs r and R , but rather choose them differently at different stages in the later arguments.

We first introduce the cutoff R alone, i.e., we bound the effect of replacing w by $w_R(x, y) = \chi_\ell(x) V_R(x - y) \chi_\ell(y)$, where $V_R(x) = |x|^{-1} e^{-|x|/R} = Y_{R^{-1}}(x)$. Thus, since $R \leq \omega(t)^{-1}\ell$, the Fourier transforms satisfy

$$\widehat{Y}_{\omega/\ell}(k) - \widehat{V}_R(k) = 4\pi \left(\frac{1}{k^2 + (\omega(t)/\ell)^2} - \frac{1}{k^2 + R^{-2}} \right) \geq 0.$$

(We use the convention that $\hat{f}(k) = \int f(x)e^{-ikx} dx$.) Hence $w(x, y) - w_R(x, y) = \chi_\ell(x) (Y_{\omega/\ell} - V_R)(x - y)\chi_\ell(y)$ defines a positive semi-definite kernel. Note, moreover, that $(Y_{\omega/\ell} - V_R)(0) = R^{-1} - \omega/\ell \leq R^{-1}$. Thus,

$$\begin{aligned} & \sum_{1 \leq i < j \leq n} w(x_i, x_j) - \rho \sum_{j=1}^n \int w(x_j, y) dy + \frac{1}{2} \rho^2 \iint w(x, y) dx dy \\ & - \left(\sum_{1 \leq i < j \leq n} w_R(x_i, x_j) - \rho \sum_{j=1}^n \int w_R(x_j, y) dy + \frac{1}{2} \rho^2 \iint w_R(x, y) dx dy \right) \\ & = \frac{1}{2} \iint \left[\sum_i^n \delta(x - x_i) - \rho \right] (w - w_R)(x, y) \left[\sum_i^n \delta(y - x_i) - \rho \right] dx dy \\ & - \frac{1}{2} \sum_i^n \chi_\ell(x_i)^2 (Y_{\omega/\ell} - V_R)(0) \geq -\frac{1}{2} n (Y_{\omega/\ell} - V_R)(0) = -\frac{1}{2} n R^{-1}. \end{aligned} \quad (12)$$

We now bound the effect of replacing w_R by $w_{r,R}$. I.e., we are replacing $V_R(x) = |x|^{-1}e^{-|x|/R}$ by $|x|^{-1}(e^{-|x|/R} - e^{-|x|/r})$. This will lower the repulsive terms and for the attractive term we get

$$\begin{aligned} -\rho \sum_{j=1}^n \int w_R(x_j, y) dy & \geq -\rho \sum_{j=1}^n \int w_{r,R}(x_j, y) dy \\ & - n \rho \sup_x \int \chi_\ell(x) \frac{e^{-|x-y|/r}}{|x-y|} \chi_\ell(y) dy \\ & \geq -\rho \sum_{j=1}^n \int w_{r,R}(x_j, y) dy - C n \rho r^2. \end{aligned} \quad (13)$$

If we combine the bounds (12) and (13) we have the following result.

Lemma 4.1 (Long and short distance potential cutoffs). *Consider the Hamiltonian*

$$\begin{aligned} H_{\ell,r,R}^n & = \sum_{j=1}^n \left(-\frac{1}{2} \gamma^{-1} \Delta_{\ell,j} - \rho \int w_{r,R}(x_j, y) dy \right) + \sum_{1 \leq i < j \leq n} w_{r,R}(x_i, x_j) \\ & + \frac{1}{2} \rho^2 \iint w_{r,R}(x, y) dx dy, \end{aligned} \quad (14)$$

where $w_{r,R}$ is given in (10) and (11) with $0 < r \leq R \leq \omega(t)^{-1}\ell$ and $-\Delta_\ell$ as before is the Neumann Laplacian for the cube $[-\ell/2, \ell/2]^3$. Then the Hamiltonian H_ℓ^n defined in (9) obeys the lower bound

$$H_\ell^n \geq H_{\ell,r,R}^n - \frac{1}{2} n R^{-1} - C_1 n \rho r^2.$$

A similar argument gives the following result.

Lemma 4.2. *With the same notation as above we have for $0 < r' \leq r \leq R \leq R' \leq \omega(t)^{-1}\ell$ that*

$$H_{\ell,r',R'}^n \geq H_{\ell,r,R}^n - \frac{1}{2} n R^{-1} - C_1 n \rho r^2.$$

Proof. Simply note that $V_{r',R'}(x) - V_{r,R}(x) = Y_{R'-1}(x) - Y_{R-1}(x) + Y_{r-1}(x) - Y_{r'-1}(x)$ and now use the same arguments as before. $\square$

Corollary 4.3 (The particle number n cannot be too small). *There exists a constant $C > 0$ such that if $\omega(t)^{-1}\rho^{1/3}\ell > C$ then $H_\ell^n \geq 0$ if $n \leq C\rho\ell^3$.*

Proof. Choose $R = \rho^{-1/3}$ and $r = \frac{1}{2}R$. Then we may assume that $R \leq \omega(t)^{-1}\ell$ since $\omega(t)^{-1}\rho^{1/3}\ell$ is large. From Lemma 4.1 we see immediately that

$$\begin{aligned} H_\ell^n &\geq -\sum_{j=1}^n \rho \int w_{r,R}(x_j, y) dy + \frac{1}{2}\rho^2 \iint w_{r,R}(x, y) dx dy - Cn\rho R^2 \\ &\geq -n \sup_x \rho \int w_{r,R}(x, y) dy + \frac{1}{2}\rho^2 \iint w_{r,R}(x, y) dx dy - Cn\rho R^2. \end{aligned}$$

The corollary follows since $\sup_x \int w_{r,R}(x, y) dy \leq 4\pi R^2$ and with the given choice of R and r it is easy to see that $\frac{1}{2}\iint w_{r,R}(x, y) dx dy \geq cR^2\ell^3$. $\square$

5. Bound on the Unimportant Part of the Hamiltonian

In this section we shall bound the Hamiltonian $H_{\ell,r,R}^n$ given in (14). We emphasize that we do not necessarily have neutrality in the cube, i.e., n and $\rho\ell^3$ may be different. We are simply looking for a lower bound to $H_{\ell,r,R}^n$, that holds for all n . The goal is to find a lower bound that will allow us to conclude that the optimal n , i.e., the value for which the energy of the Hamiltonian is smallest, is indeed close to the neutral value.

We shall express the Hamiltonian in second quantized language. This is purely for convenience. We stress that we are not in any way changing the model by doing this and the treatment is entirely rigorous and could have been done without the use of second quantization.

Let $u_p, \ell p/\pi \in (\mathbb{N} \cup \{0\})^3$ be an orthonormal basis of eigenfunctions of the Neumann Laplacian $-\Delta_\ell$ such that $-\Delta_\ell u_p = |p|^2 u_p$. I.e.,

$$u_p(x_1, x_2, x_3) = c_p \ell^{-3/2} \prod_{j=1}^3 \cos\left(\frac{p_j \pi (x_j + \ell/2)}{\ell}\right),$$

where the normalization satisfies $c_0 = 1$ and in general $1 \leq c_p \leq \sqrt{8}$. The function $u_0 = \ell^{-3/2}$ is the constant eigenfunction with eigenvalue 0. We note that for $p \neq 0$ we have

$$(u_p, -\Delta_\ell u_p) \geq \pi^2 \ell^{-2}. \quad (15)$$

We now express the Hamiltonian $H_{\ell,r,R}^n$ in terms of the creation and annihilation operators $a_p = a(u_p)$ and $a_p^* = a(u_p)^*$.

Define

$$\widehat{w}_{pq,\mu\nu} = \iint w_{r,R}(x, y) u_p(x) u_q(y) u_\mu(x) u_\nu(y) dx dy.$$

We may then express the two-body repulsive potential as

$$\sum_{1 \leq i < j \leq n} w_{r,R}(x_i, x_j) = \frac{1}{2} \sum_{pq,\mu\nu} \widehat{w}_{pq,\mu\nu} a_p^* a_q^* a_\nu a_\mu,$$

where the right-hand side is considered restricted to the n -particle subspace. Likewise the background potential can be written

$$-\rho \sum_{j=1}^n w_{r,R}(x_j, y) dy = -\rho \ell^3 \sum_{pq} \widehat{w}_{0p,0q} a_p^* a_q$$

and the background-background energy

$$\frac{1}{2} \rho^2 \iint w_{r,R}(x, y) dx dy = \frac{1}{2} \rho^2 \ell^6 \widehat{w}_{00,00}.$$

We may therefore write the Hamiltonian as

$$\begin{aligned} H_{\ell,r,R}^n &= \frac{1}{2} \gamma^{-1} \sum_p |p|^2 a_p^* a_p + \frac{1}{2} \sum_{pq,\mu\nu} \widehat{w}_{pq,\mu\nu} a_p^* a_q^* a_\nu a_\mu \\ &\quad - \rho \ell^3 \sum_{pq} \widehat{w}_{0p,0q} a_p^* a_q + \frac{1}{2} \rho^2 \ell^6 \widehat{w}_{00,00}. \end{aligned} \quad (16)$$

We also introduce the operators $\widehat{n}_0 = a_0^* a_0$ and $\widehat{n}_+ = \sum_{p \neq 0} \cdot$. These operators represent the number of particles in the condensate state created by a_0^* and the number of particle *not* in the condensate. Note that on the subspace where the total particle number is n , both of these operators are non-negative and $\widehat{n}_+ = n - \widehat{n}_0$.

Using the bounds on the long and short distance cutoffs in Lemma 4.1 we may immediately prove a simple bound on the expectation value of $\widehat{n}_+$.

Lemma 5.1 (Simple bound on the number of excited particles). *There is a constant $C > 0$ such that if $\omega(t)^{-1} \rho^{1/3} \ell > C$ then for any state such that the expectation $\langle H_\ell^n \rangle \leq 0$, the expectation of the number of excited particles satisfies $\langle \widehat{n}_+ \rangle \leq C n \rho^{-1/6} (\rho^{1/4} \ell)^2$.*

Proof. We simply choose $r = R = \rho^{-1/3}$ in Lemma 4.1. This is allowed since $R \leq \omega(t)^{-1} \ell$ is ensured from the assumption that $\omega(t)^{-1} \rho^{1/3} \ell$ is large. We then obtain

$$H_\ell^n \geq \sum_{j=1}^n -\frac{1}{2} \gamma^{-1} \Delta_{\ell,j} - \frac{1}{2} n R^{-1} - C n \rho r^2 \geq \sum_{j=1}^n -\frac{1}{2} \gamma^{-1} \Delta_{\ell,j} - C n \rho^{1/3}.$$

The bound on $\langle \widehat{n}_+ \rangle$ follows since the bound on the gap (15) implies that $\langle \sum_{j=1}^n -\Delta_{\ell,j} \rangle \geq \langle \widehat{n}_+ \rangle \pi^2 \ell^{-2}$. $\square$

Motivated by Foldy's use of the Bogolubov approximation it is our goal to reduce the Hamiltonian $H_{\ell,r,R}^n$ so that it has only what we call quadratic terms, i.e., terms which contain precisely two a_p^* with $p \neq 0$. More precisely, we want to be able to ignore all terms containing the coefficients

- $\widehat{w}_{00,00}$.
- $\widehat{w}_{p0,q0} = \widehat{w}_{0p,0q}$, where $p, q \neq 0$. These terms are in fact quadratic, but do not appear in the Foldy Hamiltonian. We shall prove that they can also be ignored.
- $\widehat{w}_{p0,00} = \widehat{w}_{0p,00} = \widehat{w}_{00,p0} = \widehat{w}_{00,0p}$, where $p \neq 0$.
- $\widehat{w}_{pq,\mu 0} = \widehat{w}_{\mu 0,pq} = \widehat{w}_{qp,0\mu} = \widehat{w}_{0\mu,qp}$, where $p, q, \mu \neq 0$.

- $\widehat{w}_{pq,\mu\nu}$, where $p, q, \mu, \nu \neq 0$. The sum of all these terms form a non-negative contribution to the Hamiltonian and can, when proving a lower bound, either be ignored or used to control error terms.

We shall consider these cases one at a time.

Lemma 5.2 (Control of terms with $\widehat{w}_{00,00}$). *The sum of the terms in $H_{\ell,r,R}^n$ containing $\widehat{w}_{00,00}$ is equal to*

$$\begin{aligned} \frac{1}{2} \widehat{w}_{00,00} \left[\left(\widehat{n}_0 - \rho \ell^3 \right)^2 - \widehat{n}_0 \right] \\ = \frac{1}{2} \widehat{w}_{00,00} \left[\left(n - \rho \ell^3 \right)^2 + (\widehat{n}_+)^2 - 2 \left(n - \rho \ell^3 \right) \widehat{n}_+ - \widehat{n}_0 \right]. \end{aligned}$$

Proof. The terms containing $\widehat{w}_{00,00}$ are

$$\frac{1}{2} \widehat{w}_{00,00} \left(a_0^* a_0^* a_0 a_0 - 2 \rho \ell^3 a_0^* a_0 + \rho^2 \ell^6 \right) = \frac{1}{2} \widehat{w}_{00,00} \left(a_0^* a_0 - \rho \ell^3 \right)^2 - \frac{1}{2} \widehat{w}_{00,00} a_0^* a_0$$

using the 0commutation relation $[a_p, a_q^*] = \delta_{p,q}$. $\square$

Lemma 5.3 (Control of terms with $\widehat{w}_{p0,q0}$). *The sum of the terms in $H_{\ell,r,R}^n$ containing $\widehat{w}_{p0,q0}$ or $\widehat{w}_{0p,0q}$ with $p, q \neq 0$ is bounded below by*

$$-4\pi [\rho - n\ell^{-3}]_+ \widehat{n}_+ R^2 - 4\pi \widehat{n}_+^2 \ell^{-3} R^2,$$

where $[t]_+ = \max\{t, 0\}$.

Proof. The terms containing $\widehat{w}_{p0,q0}$ or $\widehat{w}_{0p,0q}$ are

$$\begin{aligned} \sum_{\substack{p \neq 0 \\ q \neq 0}} \left(\frac{1}{2} \widehat{w}_{p0,q0} a_p^* a_0^* a_0 a_q + \frac{1}{2} \widehat{w}_{0p,0q} a_0^* a_p^* a_q a_0 - \rho \ell^3 \widehat{w}_{p0,q0} a_p^* a_q \right) \\ = (\widehat{n}_0 - \rho \ell^3) \sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} a_p^* a_q. \end{aligned}$$

Note that $\widehat{n}_0$ commutes with $\sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} a_p^* a_q$.

We have that

$$\widehat{w}_{p0,q0} = \ell^{-3} \int \int w_{r,R}(x, y) dy u_p(x) u_q(x) dx.$$

Hence

$$\begin{aligned} \sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} a_p^* a_q &= \ell^{-3} \int \int w_{r,R}(x, y) dy \left(\sum_{p \neq 0} u_p(x) a_p^* \right) \left(\sum_{p \neq 0} u_p(x) a_p^* \right)^* dx. \\ &\leq \ell^{-3} \sup_{x'} \int w_{r,R}(x', y) dy \int \left(\sum_{p \neq 0} u_p(x) a_p^* \right) \left(\sum_{p \neq 0} u_p(x) a_p^* \right)^* dx. \\ &= \ell^{-3} \sup_{x'} \int w_{r,R}(x', y) dy \sum_{p \neq 0} a_p^* a_p = \ell^{-3} \sup_{x'} \int w_{r,R}(x', y) dy \widehat{n}_+. \end{aligned}$$

Since

$$\sup_x \int w_{r,R}(x,y) dy \leq \int V_{r,R}(y) dy \leq 4\pi R^2$$

we obtain the operator inequality

$$0 \leq \sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} a_p^* a_q \leq 4\pi \ell^{-3} R^2 \widehat{n}_+,$$

and the lemma follows.

Before treating the last two types of terms we shall need the following result on the structure of the coefficients $\widehat{w}_{pq,\mu\nu}$.

Lemma 5.4. *For all $p', q' \in (\pi/\ell) (\mathbb{N} \cup \{0\})^3$ and $\alpha \in \mathbb{N}$ there exists $J_{p'q'}^\alpha \in \mathbb{R}$ with $J_{p'q'}^\alpha = J_{q'p'}^\alpha$, such that for all $p, q, \mu, \nu \in (\pi/\ell) (\mathbb{N} \cup \{0\})^3$ we have*

$$\widehat{w}_{pq,\mu\nu} = \sum_\alpha J_{p\mu}^\alpha J_{q\nu}^\alpha. \tag{17}$$

Moreover we have the operator inequalities

$$0 \leq \sum_{p,p' \neq 0} \widehat{w}_{pp',00} a_p^* a_{p'} = \sum_{p,p' \neq 0} \widehat{w}_{p0,0p'} a_p^* a_{p'} \leq 4\pi \ell^{-3} R^2 \widehat{n}_+ \tag{18}$$

and

$$0 \leq \sum_{p,p',m \neq 0} \widehat{w}_{pm,mp'} a_p^* a_{p'} \leq r^{-1} \widehat{n}_+.$$

Proof. The operator $\mathcal{A}$ with integral kernel $w_{r,R}(x,y)$ is a non-negative Hilbert–Schmidt operator on $L^2(\mathbb{R}^3)$ with norm less than $\sup_k \widehat{V}_{r,R}(k) \leq 4\pi R^2$. Denote the eigenvalues of $\mathcal{A}$ by λ_α , $\alpha = 1, 2, \dots$ and corresponding orthonormal eigenfunctions by φ_α . We may assume that these functions are real. The eigenvalues satisfy $0 \leq \lambda_\alpha \leq 4\pi R^2$. We then have

$$\widehat{w}_{pq,\mu\nu} = \sum_\alpha \lambda_\alpha \int u_p(x) u_\mu(x) \varphi_\alpha(x) dx \int u_q(y) u_\nu(y) \varphi_\alpha(y) dy.$$

The identity (17) thus follows with $J_{p\mu}^\alpha = \lambda_\alpha^{1/2} \int u_p(x) u_\mu(x) \varphi_\alpha(x) dx$.

If P denotes the projection onto the constant functions we may also consider the operator $(I - P)\mathcal{A}(I - P)$. Denote its eigenvalues and eigenfunctions by λ'_α and φ'_α . Then again $0 \leq \lambda'_\alpha \leq 4\pi R^2$. Hence we may write

$$\widehat{w}_{p0,0p'} = \ell^{-3} \sum_\alpha \lambda'_\alpha \int u_p(x) \varphi'_\alpha(x) dx \int u_{p'}(y) \varphi'_\alpha(y) dy.$$

Thus, since all φ'_α are orthogonal to constants we have

$$\begin{aligned} & \sum_{p, p' \neq 0} \widehat{w}_{p0,0p'} a_p^* a_{p'} \\ &= \ell^{-3} \sum_{\alpha} \lambda'_\alpha \left(\sum_{p \neq 0} \int u_p(x) \varphi'_\alpha(x) dx a_p^* \right) \left(\sum_{p \neq 0} \int u_p(x) \varphi'_\alpha(x) dx a_p^* \right)^* \\ &= \ell^{-3} \sum_{\alpha} \lambda'_\alpha a^* (\varphi'_\alpha) a (\varphi'_\alpha). \end{aligned}$$

The inequalities (18) follow immediately from this.

The fact that $\sum_{p, p', m \neq 0} \widehat{w}_{pm, mp'} a_p^* a_{p'} \geq 0$ follows from the representation (17). Moreover, since the kernel $w_{R,r}(x, y)$ is a continuous function we have that $w_{r,R}(x, x) = \sum_{\alpha} \lambda_{\alpha} \varphi_{\alpha}(x)^2$ for almost all x and hence

$$\sum_{m \neq 0} \widehat{w}_{pm, mp'} = \int u_p(x) u_{p'}(x) w_{r,R}(x, x) dx - \widehat{w}_{p0,0p'}.$$

We therefore have

$$\begin{aligned} \sum_{p, p', m \neq 0} \widehat{w}_{pm, mp'} a_p^* a_{p'} &\leq \sum_{p, p' \neq 0} \int u_p(x) u_{p'}(x) W_{r,R}(x, x) dx a_p^* a_{p'} \\ &= \int w_{r,R}(x, x) \left(\sum_{p \neq 0} u_p(x) a_p^* \right) \left(\sum_{p \neq 0} u_p(x) a_p^* \right)^* dx \\ &\leq \sup_{x'} w_{r,R}(x', x') \int \left(\sum_{p \neq 0} u_p(x) a_p^* \right) \left(\sum_{p \neq 0} u_p(x) a_p^* \right)^* dx \\ &= \sup_{x'} w_{r,R}(x', x') \widehat{n}_+ \end{aligned}$$

and the lemma follows since $\sup_{x'} w_{r,R}(x', x') \leq r^{-1}$. $\square$

Lemma 5.5 (Control of terms with $\widehat{w}_{p0,00}$). *The sum of the terms in $H_{\ell,r,R}^n$ containing $\widehat{w}_{p0,00}$, $\widehat{w}_{0p,00}$, $\widehat{w}_{00,p0}$, or $\widehat{w}_{00,0p}$, with $p \neq 0$ is, for all $\varepsilon > 0$, bounded below by*

$$-\varepsilon^{-1} 4\pi \ell^{-3} R^2 \widehat{n}_0 \widehat{n}_+ - \varepsilon \widehat{w}_{00,00} (\widehat{n}_0 + 1 - \rho \ell^3)^2, \quad (19)$$

and by

$$\begin{aligned} & \sum_{p \neq 0} \widehat{w}_{p0,00} \left((n - \rho \ell^3) a_p^* a_0 + a_0^* a_p (n - \rho \ell^3) \right) \\ & \quad - \varepsilon^{-1} 4\pi \ell^{-3} R^2 \widehat{n}_0 \widehat{n}_+ - \varepsilon \widehat{w}_{00,00} (\widehat{n}_+ - 1)^2. \quad (20) \end{aligned}$$

Proof. The terms containing $\widehat{w}_{p0,00}$, $\widehat{w}_{0p,00}$, $\widehat{w}_{00,p0}$, or $\widehat{w}_{00,0p}$ are

$$\begin{aligned} & \sum_{p \neq 0} \frac{1}{2} \widehat{w}_{p0,00} \left(2a_p^* a_0^* a_0 a_p - 2\rho \ell^3 a_0^* a_p - 2\rho \ell^3 a_p^* a_0 \right) \\ &= \sum_{p \neq 0} \widehat{w}_{p0,00} \left((\widehat{n}_0 - \rho \ell^3) a_p^* a_0 + a_0^* a_p (\widehat{n}_0 - \rho \ell^3) \right) \\ &= \sum_{\alpha} \sum_{p \neq 0} J_{p0}^{\alpha} J_{00}^{\alpha} \left(a_p^* a_0 (\widehat{n}_0 + 1 - \rho \ell^3) + (\widehat{n}_0 + 1 - \rho \ell^3) a_0^* a_p \right). \end{aligned}$$

In the last term we have used the representation (17) and the commutation relation $[\widehat{n}_0, a_0] = a_0$. For all $\varepsilon > 0$ we get that the above expression is bounded below by

$$\begin{aligned} & \varepsilon^{-1} \sum_{\alpha} \sum_{p, p' \neq 0} J_{p0}^{\alpha} J_{p'0}^{\alpha} \widehat{n}_0 a_p^* a_{p'} - \varepsilon \sum_{\alpha} (J_{00}^{\alpha})^2 (\widehat{n}_0 + 1 - \rho \ell^3)^2 \\ &= -\varepsilon^{-1} \sum_{p, p' \neq 0} \widehat{w}_{p0,0p'} \widehat{n}_0 a_p^* a_{p'} - \varepsilon \widehat{w}_{00,00} (\widehat{n}_0 + 1 - \rho \ell^3)^2. \end{aligned}$$

The bound (19) follows from (18).

The second bound (20) follows in the same way if we notice that the terms containing $\widehat{w}_{p0,00}$, $\widehat{w}_{0p,00}$, $\widehat{w}_{00,p0}$, or $\widehat{w}_{00,0p}$ may be written as

$$\begin{aligned} & \sum_{p \neq 0} \widehat{w}_{p0,00} \left((n - \rho \ell^3) a_p^* a_0 + a_0^* a_p (n - \rho \ell^3) \right) \\ &+ \sum_{\alpha} \sum_{p \neq 0} J_{p0}^{\alpha} J_{00}^{\alpha} \left(a_p^* a_0 (1 - \widehat{n}_+) + (1 - \widehat{n}_+) a_0^* a_p \right). \quad \square \end{aligned}$$

Lemma 5.6 (Control of terms with $\widehat{w}_{pq,m0}$). *The sum of the terms in $H_{\ell,r,R}^n$ containing $\widehat{w}_{pq,m0}$, $\widehat{w}_{pq,0m}$, $\widehat{w}_{p0,qm}$, or $\widehat{w}_{0p,qm}$, with $p, q, m \neq 0$ is bounded below by*

$$-\varepsilon^{-1} 4\pi \ell^{-3} R^2 \widehat{n}_0 \widehat{n}_+ - \varepsilon \widehat{n}_+ r^{-1} - \varepsilon \sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_{p'}^* a_{m'} a_p,$$

for all $\varepsilon > 0$.

Proof. The terms containing $\widehat{w}_{pq,m0}$, $\widehat{w}_{pq,0m}$, $\widehat{w}_{p0,qm}$, or $\widehat{w}_{0p,qm}$ are

$$\begin{aligned}
 & \sum_{pqm \neq 0} \widehat{w}_{pqm0} \left(a_p^* a_q^* a_m a_0 + a_0^* a_m^* a_q a_p \right) \\
 &= \sum_{\alpha} \left(\left(\sum_{q \neq 0} J_{q0}^{\alpha} a_q^* a_0 \right) \left(\sum_{pm \neq 0} J_{pm}^{\alpha} a_p^* a_m \right) \right. \\
 & \quad \left. + \left(\sum_{pm \neq 0} J_{pm}^{\alpha} a_p^* a_m \right)^* \left(\sum_{q \neq 0} J_{q0}^{\alpha} a_q^* a_0 \right)^* \right) \\
 & \geq - \sum_{\alpha} \left(\varepsilon^{-1} \left(\sum_{q \neq 0} J_{q0}^{\alpha} a_q^* a_0 \right) \left(\sum_{q \neq 0} J_{q0}^{\alpha} a_0^* a_q \right) \right. \\
 & \quad \left. + \varepsilon \left(\sum_{pm \neq 0} J_{pm}^{\alpha} a_p^* a_m \right) \left(\sum_{pm \neq 0} J_{pm}^{\alpha} a_p^* a_m \right) \right).
 \end{aligned}$$

Using that $J_{pm}^{\alpha} = J_{mp}^{\alpha}$ we may write this as

$$\begin{aligned}
 & -\varepsilon^{-1} \sum_{qq' \neq 0} \widehat{w}_{q0,0q'} a_q^* a_{q'} a_0 a_0^* - \varepsilon \sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_p a_{p'}^* a_{m'} \\
 &= -\varepsilon^{-1} \sum_{qq' \neq 0} \widehat{w}_{q0,0q'} a_q^* a_{q'} a_0 a_0^* - \varepsilon \sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_p a_{p'}^* a_{m'} \\
 & \quad - \varepsilon \sum_{p,m,m' \neq 0} \widehat{w}_{mp,pm'} a_m^* a_{m'}.
 \end{aligned}$$

The lemma now follows from Lemma 5.4. $\square$

6. Analyzing the Quadratic Hamiltonian

In this section we consider the main part of the Hamiltonian. This is the “quadratic” Hamiltonian considered by Foldy. It consists of the kinetic energy and all the terms with the coefficients $\widehat{w}_{pq,00}$, $\widehat{w}_{00,pq}$, $\widehat{w}_{p0,0q}$, and $\widehat{w}_{0p,q0}$ with $p, q \neq 0$, i.e.,

$$\begin{aligned}
 H_{\text{Foldy}} &= \frac{1}{2} \mathcal{V}^{-1} \sum_p |p|^2 a_p^* a_p \\
 & \quad + \frac{1}{2} \sum_{pq \neq 0} \widehat{w}_{pq,00} \left(a_p^* a_0^* a_q a_0 + a_0^* a_p^* a_q a_0 + a_p^* a_q^* a_0 a_0 + a_0^* a_0^* a_p a_q \right) \quad (21) \\
 &= \frac{1}{2} \mathcal{V}^{-1} \sum_p |p|^2 a_p^* a_p + \sum_{pq \neq 0} \widehat{w}_{pq,00} \left(a_p^* a_q a_0^* a_0 + \frac{1}{2} a_p^* a_q^* a_0 a_0 + \frac{1}{2} a_0^* a_0^* a_p a_q \right).
 \end{aligned}$$

In order to compute all the bounds we found it necessary to include the first term in (20) into the “quadratic” Hamiltonian. We therefore define

$$H_Q = \frac{1}{2}\gamma^{-1} \sum_p |p|^2 a_p^* a_p + \sum_{p \neq 0} \widehat{w}_{p0,00} \left((n - \rho\ell^3) a_p^* a_0 + a_0^* a_p (n - \rho\ell^3) \right) + \sum_{pq \neq 0} \widehat{w}_{pq,00} \left(a_p^* a_q a_0^* a_0 + \frac{1}{2} a_p^* a_q^* a_0 a_0 + \frac{1}{2} a_0^* a_0^* a_p a_q \right). \quad (22)$$

Note that $H_{\text{Foldy}} = H_Q$ in the neutral case $n = \rho\ell^3$. Our goal is to give a lower bound on the ground state energy of the Hamiltonian H_Q .

For the sake of convenience we first enlarge the one-particle Hilbert space $L^2([-\ell/2, \ell/2]^3)$. In fact, instead of considering the symmetric Fock space over $L^2([-\ell/2, \ell/2]^3)$ we now consider the symmetric Fock space over the one-particle Hilbert space $L^2([-\ell/2, \ell/2]^3) \oplus \mathbb{C}$. Note that the larger Fock space of course contains the original Fock space as a subspace. On the larger space we have a new pair of creation and annihilation operators that we denote $\widetilde{a}_0^*$ and $\widetilde{a}_0$. These operators merely create vectors in the $\mathbb{C}$ component of $L^2([-\ell/2, \ell/2]^3) \oplus \mathbb{C}$, and so commute with all other operators.

We shall now write

$$\widetilde{a}_p = \begin{cases} a_p, & \text{if } p \neq 0 \\ \widetilde{a}_0, & \text{if } p = 0 \end{cases} \quad \text{and} \quad \widetilde{a}_p^* = \begin{cases} a_p^*, & \text{if } p \neq 0 \\ \widetilde{a}_0^*, & \text{if } p = 0 \end{cases}. \quad (23)$$

We now define the Hamiltonian

$$\widetilde{H}_Q = \frac{1}{2}\gamma^{-1} \sum_p |p|^2 \widetilde{a}_p^* \widetilde{a}_p + \sum_p \widehat{w}_{p0,00} \left((n - \rho\ell^3) \widetilde{a}_p^* a_0 + a_0^* \widetilde{a}_p (n - \rho\ell^3) \right) + \sum_{pq} \widehat{w}_{pq,00} \left(\widetilde{a}_p^* \widetilde{a}_q a_0^* a_0 + \frac{1}{2} \widetilde{a}_p^* \widetilde{a}_q^* a_0 a_0 + \frac{1}{2} a_0^* a_0^* \widetilde{a}_p \widetilde{a}_q \right), \quad (24)$$

where we no longer restrict p, q to be different from 0. Note that for all states on the larger Fock space for which $\langle \widetilde{a}_0^* \widetilde{a}_0 \rangle = 0$ we have $\langle \widetilde{H}_Q \rangle = \langle H_Q \rangle$.

For any function $\varphi \in L^2([-\ell/2, \ell/2]^3)$ we introduce the creation operator

$$\widetilde{a}^*(\varphi) = \sum_p (u_p, \varphi) \widetilde{a}_p^*.$$

Note that the sum includes $p = 0$. the difference from $a^*(\varphi)$ is given by $\widetilde{a}^*(\varphi) - a^*(\varphi) = (u_0, \varphi) (\widetilde{a}_0^* - a_0^*)$.

Then $[\widetilde{a}(\varphi), \widetilde{a}^*(\psi)] = (\varphi, \psi)$. We have introduced the “dummy” operator $\widetilde{a}_0^*$ in order for this relation to hold. One could just as well have stayed in the old space, but then the relation above would hold only for functions orthogonal to constants.

For any $k \in \mathbb{R}^3$ denote $\chi_{\ell,k}(x) = e^{ikx} \chi_\ell(x)$ and define the operators

$$b_k^* = \widetilde{a}^*(\chi_{\ell,k}) a_0 \quad \text{and} \quad b_k = \widetilde{a}(\chi_{\ell,k}) a_0^*$$

They satisfy the commutation relations

$$[b_k, b_{k'}^*] = a_0^* a_0 (\chi_{\ell,k}, \chi_{\ell,k'}) - \widetilde{a}(\chi_{\ell,k}) \widetilde{a}^*(\chi_{\ell,k'}) \\ = a_0^* a_0 \widehat{\chi_\ell^2}(k' - k) - \widetilde{a}(\chi_{\ell,k}) \widetilde{a}^*(\chi_{\ell,k'}) \quad (25)$$

We first consider the kinetic energy part of the Hamiltonian. We shall bound it using the double commutator bound in Appendix B. First we need a well known comparison between the Neumann Laplacian and the Laplacian in the whole space.

Lemma 6.1 (Neumann resolvent is bigger than free resolvent). *Let P_ℓ denote the projection in $L^2(\mathbb{R}^3)$ that projects onto $L^2([-\ell/2, \ell/2]^3)$ (identified as a subspace). Then if $-\Delta$ denotes the Laplacian on all of $\mathbb{R}^3$ and $-\Delta_\ell$ is the Neumann Laplacian on $[-\ell/2, \ell/2]^3$ we have the operator inequality*

$$(-\Delta_\ell + a)^{-1} \geq P_\ell(-\Delta + a)^{-1} P_\ell,$$

for all $a > 0$.

Proof. It is clear that for all $f \in L^2(\mathbb{R}^3)$

$$\|P_\ell(-\Delta_\ell + a)^{1/2} P_\ell(-\Delta + a)^{-1/2} f\|^2 \leq \|f\|^2,$$

and hence

$$\|(-\Delta + a)^{-1/2} P_\ell(-\Delta_\ell + a)^{1/2} P_\ell f\|^2 \leq \|f\|^2.$$

Now simply use this with $f = (-\Delta_\ell + a)^{-1/2} u$. $\square$

Lemma 6.2 (The kinetic energy bound). *There exists a constant $C' > 0$ such that if $C't < 1$, where t is the parameter used in the definition of χ_ℓ in Sect. 3, we have*

$$\left\langle \sum_p |p|^2 \tilde{a}_p^* \tilde{a}_p \right\rangle \geq (2\pi)^{-3} (1 - C't)^2 n^{-1} \int_{\mathbb{R}^3} \frac{|k|^4}{|k|^2 + (\ell t^3)^{-2}} \langle b_k^* b_k \rangle dk$$

for all states with $\langle \tilde{a}_0^* \tilde{a}_0 \rangle = 0$ and particle number equal to n , i.e., $\left\langle \left(\sum_p a_p^* a_p \right)^2 \right\rangle = \left\langle \left(\sum_p a_p^* a_p \right)^2 \right\rangle = n^2$.

Proof. Let s , with $0 < s \leq t$, be a parameter to be chosen below. Recall that t is the parameter used in the definition of χ_ℓ in Section 3. Then since $\chi_\ell^2 + \eta_\ell^2 = 1$ we have

$$\begin{aligned} -\Delta_\ell &\geq \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} = \frac{1}{2}(\chi_\ell^2 + \eta_\ell^2) \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} + \frac{1}{2} \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} (\chi_\ell^2 + \eta_\ell^2) \\ &= \chi_\ell \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} \chi_\ell + \eta_\ell \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} \eta_\ell \\ &\quad + \left[\left[\frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}}, \chi_\ell \right], \chi_\ell \right] + \left[\left[\frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}}, \eta_\ell \right], \eta_\ell \right] \\ &\geq \chi_\ell \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} \chi_\ell + \eta_\ell \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} \eta_\ell \\ &\quad - C(\ell t)^{-2} \frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}} - C\ell^{-2} s^2 t^{-4}, \end{aligned}$$

where the last inequality follows from Lemma B.1 in Appendix B. We can now repeat this calculation to get

$$\begin{aligned} -\Delta_\ell \geq & \chi_\ell \left(\frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} - C(\ell t)^{-2} \frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}} \right) \chi_\ell \\ & + \eta_\ell \left(\frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} - C(\ell t)^{-2} \frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}} \right) \eta_\ell - C\ell^{-2}s^2t^{-4} \\ & - C(\ell t)^{-2} \left(\left[\left[\frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}}, \chi_\ell \right], \chi_\ell \right] + \left[\left[\frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}}, \eta_\ell \right], \eta_\ell \right] \right). \end{aligned}$$

If we therefore use (53) in Lemma B.1 and recall that $s \leq t$ we arrive at

$$\begin{aligned} -\Delta_\ell \geq & \chi_\ell \left(\frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} - C(\ell t)^{-2} \frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}} \right) \chi_\ell \\ & + \eta_\ell \left(\frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} - C(\ell t)^{-2} \frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}} \right) \eta_\ell - C\ell^{-2}s^2t^{-4}. \end{aligned}$$

Note that for $\alpha > 0$ we have

$$\alpha \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} - C(\ell t)^{-2} \frac{-\Delta_\ell}{-\Delta_\ell + (\ell s)^{-2}} \geq -C\alpha^{-1}s^2t^{-4}\ell^{-2}.$$

Thus if we also assume that $\alpha < 1$ we have

$$-\Delta_\ell \geq (1 - \alpha)\chi_\ell \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} \chi_\ell - C\alpha^{-1}s^2t^{-4}\ell^{-2}.$$

Thus if u is a normalized function on $L^2(\mathbb{R}^3)$ which is orthogonal to constants we have according to the bound on the gap (15) that for all $0 < \delta < 1$

$$\begin{aligned} (u, -\Delta_\ell u) \geq & (1 - \delta)(1 - \alpha) \left(u, \chi_\ell \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} \chi_\ell u \right) \\ & - C(1 - \delta)\alpha^{-1}s^2t^{-4}\ell^{-2} + \delta\pi^2\ell^{-2}. \end{aligned}$$

We choose $\alpha = \delta = C'st^{-2}$ for an appropriately large constant $C' > 0$ and assume that s and t are such that δ is less than 1. Then

$$(u, -\Delta_\ell u) \geq (1 - C'st^{-2})^2 \left(u, \chi_\ell \frac{(-\Delta_\ell)^2}{-\Delta_\ell + (\ell s)^{-2}} \chi_\ell u \right).$$

If we now use Lemma 6.1 we may write this as

$$\begin{aligned} (u, -\Delta_\ell u) \geq & (1 - C'st^{-2})^2 \left(u, \chi_\ell \Delta_\ell \frac{1}{-\Delta_\ell + (\ell s)^{-2}} \Delta_\ell \chi_\ell u \right) \\ = & (1 - C'st^{-2})^2 \left(u, \chi_\ell \frac{(-\Delta)^2}{-\Delta + (\ell s)^{-2}} \chi_\ell u \right), \end{aligned}$$

where in the last inequality we have used that $\Delta\chi = \Delta_\ell\chi$ and $\chi\Delta = \chi\Delta_\ell$.

We now choose $s = t^3$ and we may then write this inequality in second quantized form as

$$\left\langle \sum_p |p|^2 \tilde{a}_p^* \tilde{a}_p \right\rangle \geq (2\pi)^{-3} (1 - C't)^2 \int_{\mathbb{R}^3} \frac{|k|^4}{|k|^2 + (\ell t^3)^{-2}} \langle \tilde{a}^*(\chi_{\ell,k}) \tilde{a}(\chi_{\ell,k}) \rangle dk$$

using that $\langle \tilde{a}_0^* \tilde{a}_0 \rangle = 0$. Since we consider only states with particle number n the inequality still holds if we insert $n^{-1} a_0^* a_0$ as in the statement of the lemma. $\square$

With the same notation as in the above lemma we may write

$$w_{r,R}(x, y) = (2\pi)^{-3} \int \hat{V}_{r,R}(k) \chi_{\ell,k}(x) \overline{\chi_{\ell,k}(y)} dk.$$

The last two sums in the Hamiltonian (24) can therefore be written as

$$(2\pi\ell)^{-3} \int \hat{V}_{r,R}(k) \left[(n - \rho\ell^3) \ell^{-3/2} \left(\widehat{\chi}_\ell(k) b_k^* + \overline{\widehat{\chi}_\ell(k)} b_k \right) + \frac{1}{2} (b_k^* b_k + b_{-k}^* b_{-k} + b_k^* b_{-k}^* + b_k b_{-k}) \right] dk - \sum_{pq} \hat{w}_{pq,00} \tilde{a}_p^* \tilde{a}_q.$$

Note that it is important here that the potential $w_{r,R}$ contains the localization function χ_ℓ .

Thus, since $\hat{V}_{r,R}(k) = \hat{V}_{r,R}(-k)$ and $\overline{\widehat{\chi}_\ell(k)} = \widehat{\chi}_\ell(-k)$ we have for states with $\langle \tilde{a}_0^* \tilde{a}_0 \rangle = 0$ that

$$\langle \tilde{H}_Q \rangle \geq \int_{\mathbb{R}^3} \langle h_Q(k) \rangle dk - \sum_{pq} \hat{w}_{pq,00} \langle \tilde{a}_p^* \tilde{a}_q \rangle, \quad (26)$$

where

$$\begin{aligned} h_Q(k) = & \frac{(1 - C't)^2}{4(2\pi)^3 \gamma n} \frac{|k|^4}{|k|^2 + (\ell t^3)^{-2}} (b_k^* b_k + b_{-k}^* b_{-k}) \\ & + \frac{\hat{V}_{r,R}(k)}{2(2\pi\ell)^3} \left[(n - \rho\ell^3) \ell^{-3/2} \left(\widehat{\chi}_\ell(k) (b_k^* + b_{-k}) + \overline{\widehat{\chi}_\ell(k)} (b_k + b_{-k}^*) \right) \right. \\ & \left. + (b_k^* b_k + b_{-k}^* b_{-k} + b_k^* b_{-k}^* + b_k b_{-k}) \right]. \end{aligned} \quad (27)$$

Theorem 6.3 (Simple case of Bogolubov's method). *For arbitrary constants $\mathcal{A} \geq \mathcal{B} > 0$ and $\kappa \in \mathbb{C}$ we have the inequality*

$$\begin{aligned} & \mathcal{A}(b_k^* b_k + b_{-k}^* b_{-k}) + \mathcal{B}(b_k^* b_{-k}^* + b_k b_{-k}) + \kappa(b_k^* + b_{-k}) + \overline{\kappa}(b_k + b_{-k}^*) \\ & \geq -\frac{1}{2}(\mathcal{A} - \sqrt{\mathcal{A}^2 - \mathcal{B}^2})([b_k, b_k^*] + [b_{-k}, b_{-k}^*]) - \frac{2|\kappa|^2}{\mathcal{A} + \mathcal{B}}. \end{aligned}$$

Proof. We may complete the square

$$\begin{aligned} & \mathcal{A}(b_k^* b_k + b_{-k}^* b_{-k}) + \mathcal{B}(b_k^* b_{-k}^* + b_k b_{-k}) + \kappa(b_k^* + b_{-k}) + \bar{\kappa}(b_k + b_{-k}^*) \\ &= D(b_k^* + \alpha b_{-k} + a)(b_k + \alpha b_{-k}^* + \bar{a}) + D(b_{-k}^* + \alpha b_k + \bar{a})(b_{-k} + \alpha b_k^* + a) \\ &\quad - D\alpha^2([b_k, b_k^*] + [b_{-k}, b_{-k}^*]) - 2D|a|^2, \end{aligned}$$

if

$$D(1 + \alpha^2) = \mathcal{A}, \quad 2D\alpha = \mathcal{B}, \quad aD(1 + \alpha) = \kappa.$$

We choose the solution $\alpha = \mathcal{A}/\mathcal{B} - \sqrt{\mathcal{A}^2/\mathcal{B}^2 - 1}$. Hence

$$D\alpha^2 = \mathcal{B}\alpha/2 = \frac{1}{2}(\mathcal{A} - \sqrt{\mathcal{A}^2 - \mathcal{B}^2}), \quad D|a|^2 = \frac{|\kappa|^2}{D(1 + \alpha^2 + 2\alpha)} = \frac{|\kappa|^2}{\mathcal{A} + \mathcal{B}}. \quad \square$$

Usually when applying Bogolubov's method the commutator $[b_k, b_k^*]$ is a positive constant. In this case the lower bound in the theorem is actually the bottom of the spectrum of the operator. If moreover, $\mathcal{A} > \mathcal{B}$ the bottom is actually an eigenvalue. In our case the commutator $[b_k, b_k^*]$ is not a constant, but according to (25) we have

$$[b_k, b_k^*] \leq \int \chi_\ell(x)^2 dx a_0^* a_0 \leq \ell^3 a_0^* a_0. \quad (28)$$

From this and the above theorem we easily conclude the following bound.

Lemma 6.4 (Lower bound on quadratic Hamiltonian). *On the subspace with n particles we have*

$$H_Q \geq -In^{5/4}\ell^{-3/4} - \frac{1}{2} \left(n - \rho\ell^3 \right)^2 \widehat{w}_{00,00} - 4\pi n^{5/4}\ell^{-3/4}(n\ell)^{-1/4},$$

where $I = \frac{1}{2}(2\pi)^{-3} \int_{\mathbb{R}^3} f(k) - (f(k)^2 - g(k)^2)^{1/2} dk$ with

$$g(k) = 4\pi \frac{1}{k^2 + (n^{1/4}\ell^{-3/4}R)^{-2}} - 4\pi \frac{1}{k^2 + (n^{1/4}\ell^{-3/4}r)^{-2}}$$

and

$$f(k) = g(k) + \frac{1}{2}\gamma^{-1}(1 - C't)^2 \frac{|k|^4}{|k|^2 + (n^{1/4}\ell^{1/4}t)^{-2}}.$$

Proof. We consider a state with $\langle \widetilde{a}_0^* \widetilde{a}_0 \rangle = 0$. Then $\langle H_Q \rangle = \langle \widetilde{H}_Q \rangle$. We shall use (26). Note first that

$$\left\langle \sum_{pq} \widehat{w}_{pq,00} \widetilde{a}_p^* \widetilde{a}_q \right\rangle = \left\langle \sum_{p,q \neq 0} \widehat{w}_{p0,0q} a_p^* a_q \right\rangle \leq 4\pi \ell^{-3} R^2 \widehat{n}_+ \leq 4\pi \ell^{-1} n$$

by (18) and the fact that $R \leq \ell$. We may of course rewrite $\ell^{-1}n = n^{5/4}\ell^{-3/4}(n\ell)^{-1/4}$.

By Theorem 6.3, (27) and (28) we have

$$h_Q(k) \geq -(\mathcal{A}_k - \sqrt{\mathcal{A}_k^2 - \mathcal{B}_k^2})n\ell^3 - \frac{\widehat{V}_{r,R}(k)^2(n - \rho\ell^3)^2}{2(2\pi)^6\ell^9(\mathcal{A}_k + \mathcal{B}_k)} |\widehat{\chi}_\ell(k)|^2,$$

where

$$\mathcal{B}_k = \frac{\hat{V}_{r,R}(k)}{2(2\pi\ell)^3}, \quad \mathcal{A}_k = \mathcal{B}_k + \frac{(1 - C't)^2}{4(2\pi)^3\gamma n} \frac{|k|^4}{|k|^2 + (\ell t^3)^{-2}}.$$

Since $\mathcal{A}_k > \mathcal{B}_k$ we have that

$$h_Q(k) \geq -(\mathcal{A}_k - \sqrt{\mathcal{A}_k^2 - \mathcal{B}_k^2})n\ell^3 - \frac{\hat{V}_{r,R}(k)(n - \rho\ell^3)^2}{2(2\pi)^3\ell^6} |\hat{\chi}_\ell(k)|^2.$$

Note that

$$\begin{aligned} & \int \frac{\hat{V}_{r,R}(k)(n - \rho\ell^3)^2}{2(2\pi)^3\ell^6} |\hat{\chi}_\ell(k)|^2 dk \\ &= \frac{1}{2} \left(\frac{n}{\ell^3} - \rho \right)^2 \iint \chi_\ell(x) V_{r,R}(x - y) \chi_\ell(y) dx dy = \frac{1}{2} (n - \rho\ell^3)^2 \hat{w}_{00,00}. \end{aligned}$$

The lemma now follows from (26) by a simple change of variables in the k integral. $\square$

As a consequence we get the following bound for the Foldy Hamiltonian.

Corollary 6.5 (Lower bound on the Foldy Hamiltonian). *The Foldy Hamiltonian in (21) satisfies*

$$H_{\text{Foldy}} \geq -In^{5/4}\ell^{-3/4} - 4\pi n^{5/4}\ell^{-3/4}(n\ell)^{-1/4}. \quad (29)$$

There is constant $C > 0$ such that if $\rho^{1/4}R > C$, $\rho^{1/4}\ell t^3 > C$, and $t < C^{-1}$ then the Foldy Hamiltonian satisfies the bound

$$H_{\text{Foldy}} \geq \frac{1}{4} \sum_p |p|^2 a_p^* a_p - Cn^{5/4}\ell^{-3/4}. \quad (30)$$

Proof. Lemma 6.4 holds for all ρ hence also if we had replaced ρ by n/ℓ^3 in this case we get (29).

The integral I satisfies the bound

$$I \leq \frac{1}{2}(2\pi)^{-3} \int_{\mathbb{R}^3} \max \left\{ g(k), \frac{1}{2} g(k)^2 (f(k) - g(k))^{-1} \right\} dk.$$

By Corollary 4.3 we may assume that $n \geq c\rho\ell^3$. Hence I is bounded by a constant as long as $\rho^{1/4}R$ and $\rho^{1/4}\ell t^3$ are sufficiently large and t is sufficiently small (which also ensures that γ is close to 1). Note that we do not have to make any assumptions on r . Moreover, if this is true we also have that $n\ell \geq c\rho\ell^4$ is large and hence $(n\ell)^{-1}$ is small. This would give the bound in the corollary except for the first positive term. The above argument, however, also holds (with different constants) if we replace the kinetic energy in the Foldy Hamiltonian by $\frac{1}{2}(\gamma^{-1} - \frac{1}{2}) \sum_p |p|^2 a_p^* a_p$ (assuming that $\gamma < 2$). This proves the corollary. $\square$

Note that if

$$n^{1/4}\ell^{-3/4}R \rightarrow \infty, \; n^{1/4}\ell^{-3/4}r \rightarrow 0, \; n^{1/4}\ell^{1/4}t^3 \rightarrow \infty, \; \text{and } t \rightarrow 0 \tag{31}$$

it follows by dominated convergence that I converges to

$$\begin{aligned} &\frac{1}{2}(2\pi)^{-3} \int_{\mathbb{R}^3} 4\pi |k|^{-2} + \frac{1}{2}|k|^2 - \left((4\pi |k|^{-2} + \frac{1}{2}|k|^2)^2 - (4\pi |k|^{-2})^2 \right)^{1/2} dk \\ &= (2/\pi)^{3/4} \int_0^\infty 1 + x^4 - x^2 \left(x^4 + 2 \right)^{1/2} dx = - \left(\frac{4\pi}{3} \right)^{1/4} A, \end{aligned}$$

where A was given in (6). Thus if we can show that $n \sim \rho \ell^3$ we see that the term $-In^{5/4}\ell^{-3/4} \sim -I\rho^{1/4}n$ agrees with Foldy’s calculation (5) for the little box of size ℓ .

Our task is now to show that indeed $n \sim \rho \ell^3$, i.e., that we have approximate neutrality in each little box and that the term above containing the integral I is indeed the leading term.

7. Simple Bounds on n and $\widehat{n}_+$

The Lemmas 4.1, 5.2, 5.3, 5.5, and 5.6 together with Lemma 6.4 or Corollary 6.5 control all terms in the Hamiltonian H_ℓ^n except the positive term

$$\frac{1}{2} \sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_{p'}^* a_{m'} a_p.$$

If we use (30) in Corollary 6.5 together with the other bounds we obtain the following bound if $\rho^{1/4}R$ and $\rho^{1/4}\ell t^3$ are sufficiently large and t is sufficiently small

$$\begin{aligned} H_\ell^n &\geq \frac{1}{4} \sum_p |p|^2 a_p^* a_p - Cn^{5/4}\ell^{-3/4} - \frac{1}{2}nR^{-1} - Cn\rho r^2 \\ &\quad + \frac{1}{2}\widehat{w}_{00,00} \left[\left(\widehat{n}_0 - \rho \ell^3 \right)^2 - \widehat{n}_0 \right] \\ &\quad - 4\pi [\rho - n\ell^{-3}]_+ \widehat{n}_+ R^2 - 4\pi \widehat{n}_+^2 \ell^{-3} R^2 \\ &\quad - \varepsilon^{-1} 8\pi \ell^{-3} R^2 \widehat{n}_0 \widehat{n}_+ - \varepsilon \widehat{w}_{00,00} (\widehat{n}_0 + 1 - \rho \ell^3)^2 \\ &\quad - \varepsilon \widehat{n}_+ r^{-1} + \left(\frac{1}{2} - \varepsilon \right) \sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_{p'}^* a_{m'} a_p. \end{aligned}$$

The assumptions on $\rho^{1/4}R$, $\rho^{1/4}\ell t^3$, and t are needed in order to bound the integral I above by a constant. If we choose $\varepsilon = 1/4$, use $\widehat{w}_{00,00} \leq 4\pi R^2 \ell^{-3}$ and ignore the last

positive term in the bound above we arrive at

$$\begin{aligned}
 H_\ell^n &\geq \frac{1}{4} \sum_p |p|^2 a_p^* a_p - C n^{5/4} \ell^{-3/4} - \frac{1}{2} n R^{-1} - C n \rho r^2 + \frac{1}{4} \widehat{w}_{00,00} (\widehat{n}_0 - \rho \ell^3)^2 \\
 &\quad - 4\pi [\rho - n \ell^{-3}]_+ \widehat{n}_+ R^2 - 4\pi \widehat{n}_+^2 \ell^{-3} R^2 \\
 &\quad - 32\pi \ell^{-3} R^2 \widehat{n}_0 \widehat{n}_+ - 4\pi R^2 \ell^{-3} \left(\widehat{n}_0 - \frac{1}{2} \rho \ell^3 + \frac{1}{4} \right) - \frac{1}{4} \widehat{n}_+ r^{-1} \\
 &\geq \frac{1}{4} \sum_p |p|^2 a_p^* a_p - C n^{5/4} \ell^{-3/4} - \frac{1}{2} n R^{-1} - C n \rho r^2 + \frac{1}{4} \widehat{w}_{00,00} (\widehat{n}_0 - \rho \ell^3)^2 \\
 &\quad - 48\pi \ell^{-3} R^2 n \widehat{n}_+ - 4\pi R^2 \ell^{-3} \left(\widehat{n}_0 + \frac{1}{4} \right) - \frac{1}{4} \widehat{n}_+ r^{-1},
 \end{aligned} \tag{32}$$

where in the last inequality we have used that $\rho \ell^3 \leq 2n$, $\widehat{n}_0 \leq n$ and $\widehat{n}_+ \leq n$.

Lemma 7.1 (Simple bound on n). *Let $\omega(t)$ be the function described in Lemma 3.1. There is a constant $C > 0$ such that if $(\rho^{1/4} \ell) t^3 > C$ and $(\rho^{1/4} \ell) \rho^{-1/12}$, t , and $\omega(t)(\rho^{1/4} \ell)^{-1}$ are smaller than C^{-1} then for any state with $\langle H_\ell^n \rangle \leq 0$ we have $C^{-1} \rho \ell^3 \leq n \leq C \rho \ell^3$.*

Proof. The lower bound follows from Corollary 4.3. To prove the upper bound on n we choose $R = \omega(t)^{-1} \ell$ (the maximally allowed value) and $r = b \omega(t)^{-1} \ell$, where we shall choose b sufficiently small, in particular $b < 1/2$. We then have that $\rho^{1/4} R = \omega(t)^{-1} \rho^{1/4} \ell$ is large. Moreover $\widehat{w}_{00,00} \geq C R^2 \ell^{-3} = C \omega(t)^{-2} \ell^{-1}$ for some constant $C > 0$ and we get from (32) and Lemma 5.1 that

$$\begin{aligned}
 \langle H_\ell^n \rangle &\geq \ell^{-1} [-C n^{5/4} \ell^{1/4} - \frac{1}{2} n \omega(t) - C b^2 \omega(t)^{-2} n^2 + C \omega(t)^{-2} (\widehat{n}_0 - \rho \ell^3)^2 \\
 &\quad - 48\pi \omega(t)^{-2} \rho^{-1/6} (\ell \rho^{1/4})^2 n^2 - 4\pi \omega(t)^{-2} (n + \frac{1}{4}) - \frac{1}{4} n b^{-1} \omega(t)],
 \end{aligned}$$

where we have again used that $c \rho \ell^3 \leq n$, $\widehat{n}_0 \leq n$ and $\widehat{n}_+ \leq n$. Note that

$$n^{5/4} \ell^{1/4} \leq C \omega(t)^{-2} n^2 (\rho^{1/4} \ell)^{-2} \rho^{-1/4} \omega(t)^2$$

and $n \omega(t) \leq C \omega(t)^{-2} n^2 \rho^{-1} \omega(t)^3$. From Lemma 5.1 we know that $\widehat{n}_0 \geq n(1 - C \rho^{-1/6} (\ell \rho^{1/4})^2)$. By choosing b small enough we see immediately that $n \leq C \rho \ell^3$. $\square$

Using this result as an input in (32) we can get a better bound on n than above and a better bound on $\langle \widehat{n}_+ \rangle$ than given in Lemma 5.1. In particular, the next lemma in fact implies that we have near neutrality, i.e., that n is nearly $\rho \ell^3$.

Lemma 7.2 (Improved bounds on n and $\langle \widehat{n}_+ \rangle$). *There exists a constant $C > 0$ such that if $(\rho^{1/4} \ell) t^3 > C$ and $(\rho^{1/4} \ell) \rho^{-1/12}$, t , and $\omega(t)(\rho^{1/4} \ell)^{-1}$ are smaller than C^{-1} then for any state with $\langle H_\ell^n \rangle \leq 0$ we have $\langle \sum_p |p|^2 a_p^* a_p \rangle \leq C \rho^{5/4} \ell^3 (\rho^{1/4} \ell)$ and*

$$\langle \widehat{n}_+ \rangle \leq C n \rho^{-1/4} (\rho^{1/4} \ell)^3 \quad \text{and} \quad \left(\frac{n - \rho \ell^3}{\rho \ell^3} \right)^2 \leq C \rho^{-1/4} (\rho^{1/4} \ell)^3.$$

For any other state with $\langle H_{\ell, r', R'}^n \rangle \leq 0$ we have the same bound on $\langle \widehat{n}_+ \rangle'$ if $r' \leq \rho^{-3/8} (\rho^{1/4} \ell)^{1/2}$ and $R' \geq a (\rho^{1/4} \ell)^{-2} \ell$ where $a > 0$ is an appropriate constant.

Proof. Inserting the bound $n \leq C\rho\ell^3$ into (32) gives

$$\begin{aligned} H_\ell^n \geq & \frac{1}{4} \sum_p |p|^2 a_p^* a_p - C\rho^{5/4}\ell^3 - \frac{1}{2}\rho\ell^3 R^{-1} - C\rho^2\ell^3 r^2 + \frac{1}{4}\widehat{w}_{00,00}(\widehat{n}_0 - \rho\ell^3)^2 \\ & - CR^2\rho\widehat{n}_+ - CR^2\left(\rho + \frac{1}{4}\ell^{-3}\right) - \frac{1}{4}\widehat{n}_+ r^{-1}. \end{aligned}$$

We now choose $r = \rho^{-3/8}(\rho^{1/4}\ell)^{1/2}$ and $R = a(\rho^{1/4}\ell)^{-2}\ell$, where we shall choose a below, independently of ρ , $\rho^{1/4}\ell$, and t . Note that since $\omega(t)(\rho^{1/4}\ell)^{-2}$ is small we may assume that $R \leq \omega(t)^{-1}\ell$ as required and since $(\rho^{1/4}\ell)\rho^{-1/12}$ is small we may assume that $r \leq R$. Moreover $r^{-1} = \rho^{-1/8}(\rho^{1/4}\ell)^{3/2}\ell^{-2}$ and $R^2\rho = a^2(\rho^{1/4}\ell)^{-4}\ell^2\rho = a^2\ell^{-2}$. Hence, since $\sum_p |p|^2 a_p^* a_p \geq \pi^2\ell^{-2}\widehat{n}_+$ (see 15), we have

$$\begin{aligned} H_\ell^n \geq & \frac{1}{8} \sum_p |p|^2 a_p^* a_p + \left(\frac{\pi^2}{8} - a^2 - \frac{1}{4}\rho^{-1/8}(\rho^{1/4}\ell)^{3/2}\right)\ell^{-2}\widehat{n}_+ \\ & + \frac{1}{4}\widehat{w}_{00,00}(\widehat{n}_0 - \rho\ell^3)^2 \\ & - \left(\frac{1}{2a} + C\right)\rho^{5/4}\ell^3(\rho^{1/4}\ell) - Ca^2\rho^{5/4}\ell^3(\rho^{1/4}\ell)^{-5}(1 + (\rho^{1/4}\ell)^{-3}\rho^{-1/4}). \end{aligned}$$

By choosing a appropriately (independently of ρ , $\rho^{1/4}\ell$, and t) we immediately get the bound on $\langle \sum_p |p|^2 a_p^* a_p \rangle$ and the bound $\ell^{-2}\langle \widehat{n}_+ \rangle \leq C\rho^{5/4}\ell^3(\rho^{1/4}\ell)$, which implies the stated bound on $\langle \widehat{n}_+ \rangle$. The bound on $(n - \rho\ell^3)^2(\rho\ell^3)^{-2}$ follows since we also have $\widehat{w}_{00,00}(\langle \widehat{n}_0 - \rho\ell^3 \rangle^2) \leq C\rho^{5/4}\ell^3(\rho^{1/4}\ell)$ and

$$\begin{aligned} \widehat{w}_{00,00}(\langle \widehat{n}_0 - \rho\ell^3 \rangle^2) & \geq CR^2\ell^{-3}(\langle \widehat{n}_0 \rangle - \rho\ell^3)^2 \\ & \geq Ca^2(\rho^{1/4}\ell)^{-4}\ell^2\left(n - \rho\ell^3 - nC\rho^{-1/4}(\ell\rho^{1/4})^3\right)^2, \end{aligned}$$

where we have used the bound on $\langle \widehat{n}_+ \rangle$ which we have just proved.

The case when $\langle H_{\ell,r',R'}^n \rangle' \leq 0$ follows in the same way because we may everywhere replace H_ℓ^n by $H_{\ell,r',R'}^n$ and use Lemma 4.2 instead of Lemma 4.1. Note that in this case we already know the bound on n since we still assume the existence of the state such that $\langle H_\ell^n \rangle \leq 0$. $\square$

8. Localization of $\widehat{n}_+$

Note that Lemma 7.2 may be interpreted as saying that we have neutrality and condensation, in the sense that $\langle \widehat{n}_+ \rangle$ is a small fraction of n , in each little box. Although this bound on $\langle \widehat{n}_+ \rangle$ is sufficient for our purposes we still need to know that $\langle \widehat{n}_+^2 \rangle \sim \langle \widehat{n}_+ \rangle^2$. We shall however not prove this for a general state with negative energy. Instead we shall show that we may change the ground state, without changing its energy expectation significantly, in such a way that the possible $\widehat{n}_+$ values are bounded by $Cn\rho^{-1/4}(\rho^{1/4}\ell)^3$. To do this we shall use the method of localizing large matrices in Lemma A.1 of Appendix A.

We begin with any normalized n -particle wavefunction Ψ of the operator H_ℓ^n . Since Ψ is an n -particle wave function we may write $\Psi = \sum_{m=0}^n c_m \Psi_m$, where for all $m = 1, 2, \dots, n$, Ψ_m is a normalized eigenfunctions of $\widehat{n}_+$ with eigenvalue m . We may now

consider the $(n+1) \times (n+1)$ Hermitean matrix $\mathcal{A}$ with matrix elements $\mathcal{A}_{mm'} = (\Psi_m, H_{\ell,r,R}^n \Psi_{m'})$.

We shall use Lemma A.1 for this matrix and the vector $\psi = (c_0, \dots, c_n)$. We shall choose M in Lemma A.1 to be of the order of the upper bound on $\langle \hat{n}_+ \rangle$ derived in Lemma 7.2, e.g., M is the integer part of $n\rho^{-1/4}(\rho^{1/4}\ell)^3$. Recall that with the assumption in Lemma 7.2 we have $M \gg 1$. With the notation in Lemma A.1 we have $\lambda = (\psi, \mathcal{A}\psi) = (\Psi, H_{\ell,r,R}^n \Psi)$. Note also that because of the structure of $H_{\ell,r,R}^n$ we have, again with the notation in Lemma A.1, that $d_k = 0$ if $k > 3$. We conclude from Lemma A.1 that there exists a normalized wavefunction $\tilde{\Psi}$ with the property that the corresponding $\hat{n}_+$ values belong to an interval of length M and such that

$$(\Psi, H_{\ell,r,R}^n \Psi) \geq (\tilde{\Psi}, H_{\ell,r,R}^n \tilde{\Psi}) - CM^{-2}(|d_1| + |d_2|).$$

We shall discuss d_1, d_2 , which depend on Ψ , in detail below, but first we give the result on the localization of $\hat{n}_+$ that we shall use.

Lemma 8.1 (Localization of $\hat{n}_+$). *There is a constant $C > 0$ with the following property. If $(\rho^{1/4}\ell)t^3 > C$ and $(\rho^{1/4}\ell)\rho^{-1/12}$, t , and $\omega(t)(\rho^{1/4}\ell)^{-1}$ are less than C^{-1} and $r \leq \rho^{3/8}(\rho^{1/4}\ell)^{1/2}$, $R \geq C(\rho^{1/4}\ell)^{-2}\ell$, and Ψ is a normalized wavefunction such that*

$$(\Psi, H_{\ell,r,R}^n \Psi) \leq 0 \quad \text{and} \quad (\Psi, H_{\ell,r,R}^n \Psi) \leq -C(n\rho^{-1/4}(\rho^{1/4}\ell)^3)^{-2}(|d_1| + |d_2|) \quad (33)$$

then there exists a normalized wave function $\tilde{\Psi}$, which is a linear combination of eigenfunctions of $\hat{n}_+$ with eigenvalues less than $Cn\rho^{-1/4}(\rho^{1/4}\ell)^3$ only, such that

$$(\Psi, H_{\ell,r,R}^n \Psi) \geq (\tilde{\Psi}, H_{\ell,r,R}^n \tilde{\Psi}) - C(n\rho^{-1/4}(\rho^{1/4}\ell)^3)^{-2}(|d_1| + |d_2|). \quad (34)$$

Here d_1 and d_2 , depending on Ψ , are given as explained in Lemma A.1.

Proof. As explained above we choose M to be of order $n\rho^{-1/4}(\rho^{1/4}\ell)^3$. We then choose $\tilde{\Psi}$ as explained above. Then (34) holds. We also know that the possible $\hat{n}_+$ values of $\tilde{\Psi}$ range in an interval of length M . We do not know however, where this interval is located. The assumption (33) will allow us to say more about the location of the interval.

In fact, it follows from (33), (34) that $(\tilde{\Psi}, H_{\ell,r,R}^n \tilde{\Psi}) \leq 0$. It is then a consequence of Lemma 7.2 that $(\tilde{\Psi}, \hat{n}_+ \tilde{\Psi}) \leq Cn\rho^{-1/4}(\rho^{1/4}\ell)^3$. This of course establishes that the allowed $\hat{n}_+$ values are less than $C'n\rho^{-1/4}(\rho^{1/4}\ell)^3$ for some constant $C' > 0$. $\square$

Our final task in this section is to bound d_1 and d_2 . We have that $d_1 = (\Psi, H_{\ell,r,R}^n(1)\psi)$, where $H_{\ell,r,R}^n(1)$ is the part of the Hamiltonian $H_{\ell,r,R}^n$ containing all the terms with the coefficients $\tilde{w}_{pq,\mu\nu}$ for which precisely one or three indices are 0. These are the terms bounded in Lemmas 5.5 and 5.6. These lemmas are stated as one-sided bounds. It is clear from the proof that they could have been stated as two sided bounds. Alternatively we may observe that $H_{\ell,r,R}^n(1)$ is unitarily equivalent to $-H_{\ell,r,R}^n(1)$. This follows by applying the unitary transform which maps all operators a_p^* and a_p with $p \neq 0$ to $-a_p^*$ and $-a_p$. From Lemmas 5.5 and 5.6 we therefore immediately get the following bound on d_1 .

Lemma 8.2 (Control of d_1). *With the notation above we have for all $\varepsilon > 0$*

$$|d_1| \leq \varepsilon^{-1} 8\pi \ell^{-3} R^2 (\Psi, \hat{n}_0 \hat{n}_+ \Psi) + \varepsilon \left(\Psi, \left(\hat{n}_+ r^{-1} + \hat{w}_{00,00} (\hat{n}_0 + 1 - \rho \ell^3)^2 \right) \Psi \right) \\ + \varepsilon \left(\Psi, \sum_{p,m,p',m' \neq 0} \hat{w}_{mp',pm'} a_m^* a_{p'}^* a_{m'} a_p \Psi \right).$$

Likewise, we have that $d_2 = (\Psi, H_{\ell,r,R}^n(2)\psi)$, where $H_{\ell,r,R}^n(2)$ is the part of the Hamiltonian $H_{\ell,r,R}^n$ containing all the terms with precisely two a_0 or two a_0^* . i.e., these are the terms in the Foldy Hamiltonian, which do not commute with $\hat{n}_+$.

Lemma 8.3 (Control of d_2). *There exists a constant $C > 0$ such that if $(\rho^{1/4} \ell) t^3 > C$ and $(\rho^{1/4} \ell) \rho^{-1/12}$, t , and $\omega(t)(\rho^{1/4} \ell)^{-1}$ are less than C^{-1} and Ψ is a wave function with $(\Psi, H_{\ell}^n \Psi) \leq 0$ then with the notation above we have*

$$|d_2| \leq C \rho^{5/4} \ell^3 (\rho^{1/4} \ell) + 4\pi \ell^{-3} R^2 (\Psi, \hat{n}_+ \hat{n}_0 \Psi).$$

Proof. If we replace all the operators a_p^* and a_p with $p \neq 0$ in the Foldy Hamiltonian by $-ia_p^*$ and ia_p we get a unitarily equivalent operator. This operator however differs from the Hamiltonian H_{Foldy} only by a change of sign on the part that we denoted $H_{\ell,r,R}^n(2)$. Since both operators satisfy the bound in Corollary 6.5 we conclude that

$$|d_2| \leq \left(\Psi, \left[\frac{1}{2} \gamma^{-1} \sum_p |p|^2 a_p^* a_p + \frac{1}{2} \sum_{pq \neq 0} \hat{w}_{pq,00} (a_p^* a_0^* a_0 a_q + a_0^* a_p^* a_q a_0) \right] \Psi \right) \\ + C n^{5/4} \ell^{-3/4}.$$

Note that both sums above define positive operators. This is trivial for the first sum. For the second it follows from (18) in Lemma 5.4 since $a_0^* a_0$ commutes with all a_p^* and a_p with $p \neq 0$. The lemma now follows from (18) and from Lemma 7.2. $\square$

9. Proof of Foldy's Law

We first prove Foldy's law in a small cube. Let Ψ be a normalized n -particle wave function. We shall prove that with an appropriate choice of ℓ

$$(\Psi, H_{\ell}^n \Psi) \geq \left(\frac{4\pi}{3} \right)^{1/3} A \rho \ell^3 \left(\rho^{1/4} + o\left(\rho^{1/4}\right) \right), \quad (35)$$

where A is given in (6). Note that $A < 0$. It then follows from Lemma 3.3 that

$$E_0 \geq (1 + L/\ell)^3 \gamma \left(\frac{4\pi}{3} \right)^{1/3} A \rho \ell^3 \left(\rho^{1/4} + o\left(\rho^{1/4}\right) \right) - C(L/\ell)^2 \rho^2 \ell^5 - \frac{\omega(t)N}{2\ell}.$$

Thus, since $N = \rho L^3$ we have

$$\lim_{L \rightarrow \infty} \frac{E_0}{N} \geq \gamma \left(\frac{4\pi}{3} \right)^{1/3} A \left(\rho^{1/4} + o\left(\rho^{1/4}\right) \right) - C \rho^{1/4} \omega(t) \left(\rho^{1/4} \ell \right)^{-1}.$$

Foldy's law (5) follows since we shall choose (see below) t and ℓ in such a way that as $\rho \rightarrow \infty$ we have $t \rightarrow 0$ and hence $\gamma \rightarrow 1$ and $\omega(t)(\rho^{1/4} \ell)^{-1} \rightarrow 0$ (see condition (41) below).

It remains to prove (35). First we fix the long and short distance potential cutoffs

$$R = \omega(t)^{-1}\ell, \quad \text{and} \quad r = \rho^{-3/8}(\rho^{1/4}\ell)^{-1/2}. \quad (36)$$

We may of course assume that $(\Psi, H_\ell^n \Psi) \leq 0$. Thus n satisfies the bound in Lemma 7.2. We proceed in two steps. In Lemma 9.1 Foldy's law in the small boxes is proved under the restrictive assumption given in (37) below. Finally, in Theorem 9.2 Foldy's law in the small boxes is proved by considering the alternative case that (37) fails. Let us note that, logically speaking, this could have been done in the reverse order. I.e., we could, instead, have begun with the case that (37) fails. At the end of the section we combine Theorem 9.2 with Lemma 3.3 to show that Foldy's law in the small box implies Foldy's law Theorem 1.1.

At the end of this section we show how to choose ℓ and t so that Theorem 9.2 implies (35) and hence Theorem 1.1, as explained above.

Lemma 9.1 (Foldy's law for H_ℓ^n : restricted version). *Let R and r be given by (36). There exists a constant $C > 0$ such that if $(\rho^{1/4}\ell)t^3 > C$ and $(\rho^{1/4}\ell)\rho^{-1/12}$, t , and $\omega(t)(\rho^{1/4}\ell)^{-1}$ are less than C^{-1} then, whenever*

$$n\ell^{-3}R^2(\Psi, \hat{n}_+\Psi) \leq C^{-1}\left(\Psi, \left(\hat{w}_{00,00}(\hat{n}_0 - \rho\ell^3)^2 + \sum_{p,m,p',m' \neq 0} \hat{w}_{mp',pm'}a_m^*a_{p'}^*a_{m'}a_p\right)\Psi\right), \quad (37)$$

we have that

$$(\Psi, H_\ell^n \Psi) \geq -In^{5/4}\ell^{-3/4} - C\rho^{5/4}\ell^3 \left(\omega(t)(\rho^{1/4}\ell)^{-1} + \omega(t)^{-2}\rho^{-1/8}(\rho^{1/4}\ell)^{13/2} + \rho^{-1/8}(\rho^{1/4}\ell)^{7/2} \right),$$

with I as in Lemma 6.4.

Proof. We assume $(\Psi, H_\ell^n \Psi) \leq 0$. We proceed as in the beginning of Sect. 7, but we now use (29) of Corollary 6.5 instead of (30). We then get

$$\begin{aligned} H_\ell^n &\geq -In^{5/4}\ell^{-3/4} - 4\pi n^{5/4}\ell^{-3/4}(n\ell)^{-1/4} - \frac{1}{2}nR^{-1} - Cn\rho r^2 \\ &\quad + \frac{1}{2}\hat{w}_{00,00}\left[\left(\hat{n}_0 - \rho\ell^3\right)^2 - \hat{n}_0\right] \\ &\quad - 4\pi[\rho - n\ell^{-3}]_+\hat{n}_+R^2 - 4\pi\hat{n}_+^2\ell^{-3}R^2 \\ &\quad - \varepsilon^{-1}8\pi\ell^{-3}R^2\hat{n}_0\hat{n}_+ - \varepsilon\hat{w}_{00,00}(\hat{n}_0 + 1 - \rho\ell^3)^2 \\ &\quad - \varepsilon\hat{n}_+r^{-1} + \left(\frac{1}{2} - \varepsilon\right) \sum_{p,m,p',m' \neq 0} \hat{w}_{mp',pm'}a_m^*a_{p'}^*a_{m'}a_p. \end{aligned}$$

If we now use the assumption (37) and the facts that $\hat{n}_+ \leq n$, $\hat{n}_0 \leq n$, and $\hat{w}_{00,00} \leq 4\pi R^2\ell^{-3}$ we see with appropriate choices of ε and C that

$$\begin{aligned} H_\ell^n &\geq -In^{5/4}\ell^{-3/4} - 4\pi n^{5/4}\ell^{-3/4}(n\ell)^{-1/4} - \frac{1}{2}nR^{-1} - Cn\rho r^2 - CR^2\ell^{-3}(n+1) \\ &\quad - CR^2\ell^{-3}|n - \rho\ell^3|(\hat{n}_+ + 1) - C\hat{n}_+r^{-1}. \end{aligned}$$

If we finally insert the choices of R and r and use Lemma 7.2 we arrive at the bound in the lemma. $\square$

Theorem 9.2 (Foldy's law for H_ℓ^n). *There exists a $C > 0$ such that if $(\rho^{1/4}\ell)t^3 > C$ and $(\rho^{1/4}\ell)\rho^{-1/12}$, t , and $\omega(t)(\rho^{1/4}\ell)^{-1}$ are less than C^{-1} then for any normalized n -particle wave function Ψ we have*

$$(\Psi, H_\ell^n \Psi) \geq -In^{5/4}\ell^{-3/4} - C\rho^{5/4}\ell^3 \left(\omega(t)(\rho^{1/4}\ell)^{-1} + \omega(t)^{-1}\rho^{-1/16}(\rho^{1/4}\ell)^{29/4} + \rho^{-1/8}(\rho^{1/4}\ell)^{7/2} \right), \quad (38)$$

where I is defined in Lemma 6.4 with r and R as in (36).

Proof. According to Lemma 9.1 we may assume that

$$n\ell^{-3}R^2(\Psi, \hat{n}_+\Psi) \geq C^{-1} \left(\Psi, \left(\hat{w}_{00,00}(\hat{n}_0 - \rho\ell^3)^2 + \sum_{p,m,p',m' \neq 0} \hat{w}_{mp',pm'} a_m^* a_p^* a_{m'} a_p \right) \Psi \right), \quad (39)$$

where C is at least as big as the constant in Lemma 9.1. We still assume that $(\Psi, H_\ell^n \Psi) \leq 0$.

We begin by bounding d_1 and d_2 using Lemmas 8.2 and 8.3. We have from Lemmas 7.2 and 8.3 that

$$\begin{aligned} |d_2| &\leq C\rho^{5/4}\ell^3(\rho^{1/4}\ell) + C\ell^{-1}\omega(t)^{-2}n^2\rho^{-1/4}(\rho^{1/4}\ell)^3 \\ &\leq C[n\rho^{-1/4}(\rho^{1/4}\ell)^3]^2\rho^{5/4}\ell^3 \left((\rho^{1/4}\ell)^{-11} + \omega(t)^{-2}(\rho^{1/4}\ell)^{-7} \right) \\ &\leq C[n\rho^{-1/4}(\rho^{1/4}\ell)^3]^2\rho^{5/4}\ell^3\omega(t)^{-2}(\rho^{1/4}\ell)^{-7}. \end{aligned}$$

In order to bound d_1 we shall use (39). Together with Lemma 8.2 this gives (choosing $\varepsilon = 1/2$ say)

$$|d_1| \leq C\ell^{-3}R^2n(\Psi, \hat{n}_+\Psi) + \frac{1}{2} \left(\Psi, \left(\hat{n}_+r^{-1} + \hat{w}_{00,00}(n - \rho\ell^3 + 1) \right) \Psi \right).$$

Inserting the choices for r and R and using Lemma 7.2 gives

$$|d_1| \leq C[n\rho^{-1/4}(\rho^{1/4}\ell)^3]^2\rho^{5/4}\ell^3 \left(\omega(t)^{-2}(\rho^{1/4}\ell)^{-7} + \rho^{-1/8}(\rho^{1/4}\ell)^{-17/2} \right),$$

where we have also used that we may assume that $\rho^{-1/8}(\rho^{1/4}\ell)^{-9/2}$ is small. The assumption (33) now reads

$$(\Psi, H_{\ell,r,R}^n \Psi) \leq -C\rho^{5/4}\ell^3 \left(\omega(t)^{-2}(\rho^{1/4}\ell)^{-7} + \rho^{-1/8}(\rho^{1/4}\ell)^{-17/2} \right).$$

If this is not satisfied we see immediately that the bound (38) holds.

Thus from Lemma 8.1 it follows that we can find a normalized n -particle wavefunction $\tilde{\Psi}$ with

$$(\tilde{\Psi}, \hat{n}_+\tilde{\Psi}) \leq Cn\rho^{-1/4}(\rho^{1/4}\ell)^3 \quad \text{and} \quad (\tilde{\Psi}, \hat{n}_+^2\tilde{\Psi}) \leq Cn^2\rho^{-1/2}(\rho^{1/4}\ell)^6 \quad (40)$$

such that

$$(\Psi, H_{\ell,r,R}^n \Psi) \geq (\tilde{\Psi}, H_{\ell,r,R}^n \tilde{\Psi}) - C\rho^{5/4}\ell^3 \left(\omega(t)^{-2}(\rho^{1/4}\ell)^{-7} + \rho^{-1/8}(\rho^{1/4}\ell)^{-17/2} \right).$$

In order to analyze $(\tilde{\Psi}, H_{\ell,r,R}^n \tilde{\Psi})$ we proceed as in the beginning of Sect. 7. This time we use Lemmas 4.1, 5.2, 5.3, 5.5, and 5.6 together with Lemma 6.4 instead of Corollary 6.5. We obtain

$$\begin{aligned} H_{\ell,r,R}^n &\geq \frac{1}{2} \widehat{w}_{00,00} \left[\left(n - \rho \ell^3 \right)^2 + (\widehat{n}_+)^2 - 2 \left(n - \rho \ell^3 \right) \widehat{n}_+ - \widehat{n}_0 \right] \\ &\quad - 4\pi [\rho - n \ell^{-3}]_+ \widehat{n}_+ R^2 - 4\pi \widehat{n}_+^2 \ell^{-3} R^2 - \varepsilon \widehat{n}_+ r^{-1} - \varepsilon^{-1} 8\pi \ell^{-3} R^2 \widehat{n}_0 \widehat{n}_+ \\ &\quad - \varepsilon \widehat{w}_{00,00} (\widehat{n}_+ - 1)^2 + \left(\frac{1}{2} - \varepsilon \right) \sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_{p'}^* a_{m'} a_p \\ &\quad - \frac{1}{2} \left(n - \rho \ell^3 \right)^2 \widehat{w}_{00,00} - 4\pi n^{5/4} \ell^{-3/4} (n \ell)^{-1/4} - I n^{5/4} \ell^{-3/4}. \end{aligned}$$

This time we shall however not choose ε small, but rather big. Note that since $w_{r,R}(x, y) \leq r^{-1}$ we have $\sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_{p'}^* a_{m'} a_p \leq r^{-1} \widehat{n}_+ (\widehat{n}_+ - 1)$, which follows immediately from

$$\begin{aligned} &\sum_{p,m,p',m' \neq 0} \widehat{w}_{mp',pm'} a_m^* a_{p'}^* a_{m'} a_p \\ &= \iint w_{r,R}(x, y) \left(\sum_{p,m \neq 0} u_m(x) u_p(y) a_m a_p \right)^* \sum_{p,m \neq 0} u_m(x) u_p(y) a_m a_p dx dy. \end{aligned}$$

We therefore have

$$\begin{aligned} H_{\ell,r,R}^n &\geq -I n^{5/4} \ell^{-3/4} - 4\pi n^{5/4} \ell^{-3/4} (n \ell)^{-1/4} - C R^2 \ell^{-3} \widehat{n}_0 \\ &\quad - C \ell^{-3} R^2 |\rho \ell^3 - n| \widehat{n}_+ - 4\pi \widehat{n}_+^2 \ell^{-3} R^2 - \varepsilon \widehat{n}_+ r^{-1} - \varepsilon^{-1} 8\pi \ell^{-3} R^2 \widehat{n}_0 \widehat{n}_+ \\ &\quad - \varepsilon C R^2 \ell^{-3} \widehat{n}_+^2 - \varepsilon \widehat{n}_+^2 r^{-1}. \end{aligned}$$

If we now insert the choices of r and R , take the expectation in the state given by $\tilde{\Psi}$, and use (40) and the bound on n from Lemma 7.2 we arrive at

$$\begin{aligned} (\tilde{\Psi}, H_{\ell,r,R}^n \tilde{\Psi}) &\geq -I n^{5/4} \ell^{-3/4} - C \rho^{5/4} \ell^3 \left[(\rho^{1/4} \ell)^{-1} + \omega(t)^{-2} (\rho^{1/4} \ell)^{-1} \right. \\ &\quad + \omega(t)^{-2} \rho^{-1/8} (\rho^{1/4} \ell)^{11/2} + \omega(t)^{-2} \rho^{-1/4} (\rho^{1/4} \ell)^8 + \varepsilon \rho^{-1/8} (\rho^{1/4} \ell)^{7/2} \\ &\quad \left. + \varepsilon^{-1} \omega(t)^{-2} (\rho^{1/4} \ell)^5 + \varepsilon \omega(t)^{-2} \rho^{-1/4} (\rho^{1/4} \ell)^8 + \varepsilon \rho^{-1/8} (\rho^{1/4} \ell)^{19/2} \right]. \end{aligned}$$

If we now choose $\varepsilon = \omega(t)^{-1} \rho^{1/16} (\rho^{1/4} \ell)^{-9/4}$ we arrive at (38). $\square$

Completion of the proof of Foldy's law, Theorem 1.1. We have accumulated various errors and we want to show that they can all be made small. There are basically two parameters that can be adjusted, ℓ and t . Instead of ℓ it is convenient to use $X = \rho^{1/4} \ell$. We shall choose X as a function of ρ such that $X \rightarrow \infty$ as $\rho \rightarrow \infty$. From Lemma 7.1 we know that for some fixed $C > 0$ $C^{-1} \rho \ell^3 \leq n \leq C \rho \ell^3$. Hence according to (31) with r and R

given in (36) we have that $I \rightarrow -\left(\frac{4\pi}{3}\right)^{1/3} A$ as $\rho \rightarrow \infty$ if

$$\omega(t)^{-1} X \rightarrow \infty, \quad (41)$$

$$\rho^{1/4} X \rightarrow \infty, \quad (42)$$

$$t^3 X \rightarrow \infty, \quad (43)$$

$$t \rightarrow 0. \quad (44)$$

The hypotheses of Theorem 9.2 are valid if (41), (43), (44), and

$$\rho^{-1/12} X \rightarrow 0 \quad (45)$$

hold. From Lemma 7.2, for which the hypotheses are now automatically satisfied, we have that $n = \rho \ell^3 (1 + O(\rho^{-1/8} X^{3/2}))$ and from (45) we see that n is $\rho \ell^3$ to leading order.

With these conditions we find that the first term on the right side of (38) is, in the limit $\rho \rightarrow \infty$, exactly Foldy's law. The conditions that the other terms in (38) are of lower order are

$$(X/\omega(t))^{4/25} \rho^{-1/100} X \rightarrow 0, \quad (46)$$

$$\rho^{-1/28} X \rightarrow 0 \quad (47)$$

together with (41).

It remains to show that we can satisfy the conditions (41–47). Condition (42) is trivially satisfied since both ρ and X tend to infinity. Since $\omega(t) \sim t^{-4}$ for small t we see that (43) is implied by (41). Condition (45) is implied by (47), which is in turn implied by (41) and (46). The remaining two conditions (41) and (46) are easily satisfied by an appropriate choice of X and t as functions for ρ with $X \rightarrow \infty$ and $t \rightarrow 0$ as $\rho \rightarrow \infty$. In fact, we simply need $\rho^{1/116} t^{-16/29} \gg X \gg t^{-4}$.

The bound (35) has now been established. Hence Foldy's law Theorem 1.1 follows as discussed in the beginning of the section.

Appendix

A. Localization of Large Matrices

The following theorem allows us to reduce a big Hermitean matrix, $\mathcal{A}$, to a smaller principal submatrix without changing the lowest eigenvalue very much. (The k^{th} supra- (resp. infra-) diagonal of a matrix $\mathcal{A}$ is the submatrix consisting of all elements $a_{i,i+k}$ (resp. $a_{i+k,i}$).)

Theorem A.1 (Localization of large matrices). *Suppose that $\mathcal{A}$ is an $N \times N$ Hermitean matrix and let $\mathcal{A}^k$, with $k = 0, 1, \dots, N-1$, denote the matrix consisting of the k^{th} supra- and infra-diagonal of $\mathcal{A}$. Let $\psi \in \mathbb{C}^N$ be a normalized vector and set $d_k = (\psi, \mathcal{A}^k \psi)$ and $\lambda = (\psi, \mathcal{A} \psi) = \sum_{k=0}^{N-1} d_k$. (ψ need not be an eigenvector of $\mathcal{A}$.)*

Choose some positive integer $M \leq N$. Then, with M fixed, there is some $n \in [0, N-M]$ and some normalized vector $\phi \in \mathbb{C}^N$ with the property that $\phi_j = 0$ unless $n+1 \leq j \leq n+M$ (i.e., ϕ has length M) and such that

$$(\phi, \mathcal{A} \phi) \leq \lambda + \frac{C}{M^2} \sum_{k=1}^{M-1} k^2 |d_k| + C \sum_{k=M}^{N-1} |d_k|, \quad (48)$$

where $C > 0$ is a universal constant. (Note that the first sum starts with $k = 1$.)

Proof. It is convenient to extend the matrix $\mathcal{A}_{i,j}$ to all $-\infty < i, j < +\infty$ by defining $\mathcal{A}_{i,j} = 0$ unless $1 \leq i, j \leq N$. Similarly, we extend the vector ψ and we define the numbers d_k and the matrix $\mathcal{A}^k$ to be zero when $k \notin [0, N-1]$. We shall give the construction for M odd, the M even case being similar.

For $s \in \mathbf{Z}$ set $f(s) = A_M[M+1-2|s|]$ if $2|s| < M$ and $f(s) = 0$ otherwise. Thus, $f(s) \neq 0$ for precisely M values of s . Also, $f(s) = f(-s)$. A_M is chosen so that $\sum_s f(s)^2 = 1$.

For each $m \in \mathbf{Z}$ define the vector $\phi^{(m)}$ by $\phi_j^{(m)} = f(j-m)\psi_j$. We then define $K^{(m)} = (\phi^{(m)}, \mathcal{A}\phi^{(m)}) - (\lambda + \sigma)(\phi^{(m)}, \phi^{(m)})$. (The number σ will be chosen later.) After this, we define $K = \sum_m K^{(m)}$. Using the fact that $\sum_s f(s)^2 = 1$, we have that

$$\begin{aligned} \sum_m (\phi^{(m)}, \mathcal{A}\phi^{(m)}) &= \sum_m \sum_{k=0} (\phi^{(m)}, \mathcal{A}^k \phi^{(m)}) = \sum_s \sum_k f(s)f(k+s)(\psi, \mathcal{A}^k \psi) \\ &= \sum_s \sum_{k=0} f(s)f(k+s)d_k \end{aligned}$$

and

$$\lambda = \lambda \sum_m (\phi^{(m)}, \phi^{(m)}) = \sum_s \sum_{k=0} f(s)^2(\psi, \mathcal{A}^k \psi) = \sum_s \sum_k f(s)^2 d_k \quad (49)$$

Hence

$$K = \sum_m K^{(m)} = -\sigma - \sum_{k=1}^{N-1} d_k \gamma_k \quad (50)$$

with

$$\gamma_k = \frac{1}{2} \sum_s [f(s) - f(s+k)]^2. \quad (51)$$

Let us choose $\sigma = -\sum_{k=1}^{N-1} d_k \gamma_k$. Then, $\sum_m K^{(m)} = 0$. Recalling that not all of the $\phi^{(m)}$ equal zero, we conclude that there is at least one value of m such that (i) $\phi^{(m)} \neq 0$ and (ii) $(\phi^{(m)}, \mathcal{A}\phi^{(m)}) \leq (\lambda + \sigma)(\phi^{(m)}, \phi^{(m)})$.

This concludes the proof of (48) except for showing that $\gamma_k \leq C \frac{k^2}{k^2 + M^2}$ for all M and k . This is evident from the easily computable large M asymptotics in (51). $\square$

B. A Double Commutator Bound

Lemma B.1. *Let $-\Delta_N$ be the Neumann Laplacian of some bounded open set $\mathcal{O}$. Given $\theta \in C^\infty(\overline{\mathcal{O}})$ with $\text{supp } |\nabla \theta| \subset \mathcal{O}$ satisfying $\|\partial_i \theta\| \leq Ct^{-1}$, $\|\partial_i \partial_j \theta\| \leq Ct^{-2}$, $\|\partial_i \partial_j \partial_k \theta\| \leq Ct^{-3}$, for some $0 < t$ and all $i, j, k = 1, 2, 3$. Then for all $s > 0$ we have the operator inequality*

$$\left[\left[\frac{(-\Delta_N)^2}{-\Delta_N + s^{-2}}, \theta \right], \theta \right] \geq -Ct^{-2} \frac{-\Delta_N}{-\Delta_N + s^{-2}} - Cs^2 t^{-4}. \quad (52)$$

We also have the norm bound

$$\left\| \left[\left[\frac{-\Delta_N}{-\Delta_N + s^{-2}}, \theta \right], \theta \right] \right\| \leq C(s^2 t^{-2} + s^4 t^{-4}). \quad (53)$$

Proof. We calculate the commutator

$$\left[\frac{(-\Delta_N)^2}{-\Delta_N + s^{-2}}, \theta \right] = s^{-2} \frac{1}{-\Delta_N + s^{-2}} [-\Delta_N, \theta] \frac{1}{-\Delta_N + s^{-2}} (-\Delta_N) \\ + \frac{-\Delta_N}{-\Delta_N + s^{-2}} [-\Delta_N, \theta].$$

Likewise we calculate the double commutator

$$\left[\left[\frac{(-\Delta_N)^2}{-\Delta_N + s^{-2}}, \theta \right], \theta \right] = - \frac{-\Delta_N}{-\Delta_N + s^{-2}} [[-\Delta_N, \theta] \theta] \frac{-\Delta_N}{-\Delta_N + s^{-2}} \\ + [[-\Delta_N, \theta] \theta] \frac{-\Delta_N}{-\Delta_N + s^{-2}} + \frac{-\Delta_N}{-\Delta_N + s^{-2}} [[-\Delta_N, \theta] \theta] \\ - 2s^{-4} \frac{1}{-\Delta_N + s^{-2}} [-\Delta_N, \theta] \frac{1}{-\Delta_N + s^{-2}} [\theta, -\Delta_N] \frac{1}{-\Delta_N + s^{-2}}. \quad (54)$$

Note that $[[-\Delta_N, \theta] \theta] = -2(\nabla\theta)^2$ and thus the first term above is positive.

We claim that

$$[-\Delta_N, \theta][\theta, -\Delta_N] \leq -Ct^{-2}\Delta_N + Ct^{-4}. \quad (55)$$

To see this we simply calculate

$$[-\Delta_N, \theta][\theta, -\Delta_N] = - \sum_{i,j}^3 \left(4\partial_i(\partial_i\theta)(\partial_j\theta)\partial_j + (\partial_i^2\theta)(\partial_j^2\theta) + 2(\partial_i\theta)(\partial_i\partial_j^2\theta) \right)$$

The last two terms are bounded by Ct^{-4} . For the first term we have by the Cauchy-Schwarz inequality for operators, $BA^* + AB^* \leq \varepsilon^{-1}AA^* + \varepsilon BB^*$, for all $\varepsilon > 0$, that

$$- \sum_{i,j}^3 \partial_i(\partial_i\theta)(\partial_j\theta)\partial_j = \sum_{i,j}^3 (\partial_i(\partial_i\theta)) (\partial_j(\partial_j\theta))^* \leq -3 \sum_i^3 \partial_i(\partial_i\theta)(\partial_i\theta)\partial_i$$

and this is bounded above by $-3t^{-2}\Delta_N$ and we get (55). Inserting (55) into (54), recalling that the first term is positive, we obtain

$$\left[\left[\frac{(-\Delta_N)^2}{-\Delta_N + s^{-2}}, \theta \right], \theta \right] \geq -2(\nabla\theta)^2 \frac{-\Delta_N}{-\Delta_N + s^{-2}} - 2 \frac{-\Delta_N}{-\Delta_N + s^{-2}} (\nabla\theta)^2 \\ - Ct^{-2} \frac{-\Delta_N}{-\Delta_N + s^{-2}} - Cs^2t^{-4}.$$

Again using the Cauchy-Schwarz inequality, we have

$$2(\nabla\theta)^2 \frac{-\Delta_N}{-\Delta_N + s^{-2}} + 2 \frac{-\Delta_N}{-\Delta_N + s^{-2}} (\nabla\theta)^2 \\ \leq 2t^{-2} \left(\frac{-\Delta_N}{-\Delta_N + s^{-2}} \right)^{1/2} (\nabla\theta)^4 \left(\frac{-\Delta_N}{-\Delta_N + s^{-2}} \right)^{1/2} + 2t^{-2} \left(\frac{-\Delta_N}{-\Delta_N + s^{-2}} \right) \\ \leq Ct^{-2} \frac{-\Delta_N}{-\Delta_N + s^{-2}},$$

and (52) follows.

The bound (53) is proved in the same way. Indeed,

$$\begin{aligned} \left[\left[\frac{-\Delta_N}{-\Delta_N + s^{-2}}, \theta \right], \theta \right] &= -s^{-2} \frac{1}{-\Delta_N + s^{-2}} [[-\Delta_N, \theta], \theta] \frac{1}{-\Delta_N + s^{-2}} \\ &\quad + 2s^{-2} \frac{1}{-\Delta_N + s^{-2}} [-\Delta_N, \theta] \frac{1}{-\Delta_N + s^{-2}} [\theta - \Delta_N] \frac{1}{-\Delta_N + s^{-2}}, \end{aligned}$$

and (53) follows from $[-\Delta_N, \theta] \theta = -2(\nabla \theta)^2$ and (55). $\square$

References

- [B] Bogolubov, N.N.: J. Phys. (U.S.S.R.) **11**, 23 (1947); Bogolubov, N.N. and Zubarev, D.N.: Sov. Phys. JETP **1**, 83 (1955)
- [CLY] Conlon, J.G., Lieb, E.H. and Yau, H-T.: The $N^{7/5}$ law for charged bosons. Commun. Math. Phys. **116**, 417–448 (1988)
- [D] Dyson, F.J.: Ground-state energy of a finite system of charged particles. J. Math. Phys. **8**, 1538–1545 (1967)
- [F] Foldy, L.L.: Charged boson gas. Phys. Rev. **124**, 649–651 (1961); Errata. ibid **125**, 2208 (1962)
- [GM] Girardeau, M.: Ground state of the charged Bose gas. Phys. Rev. **127**, 1809–1818 (1962)
- [GA] Girardeau, M. and Arnowitt, R.: Theory of many-boson systems: Pair theory. Phys. Rev. **113**, 755–761 (1959)
- [GG] Graf, G.M.: Stability of matter through an electrostatic inequality. Helv. Phys. Acta **70**, 72–79 (1997)
- [KLS] Kennedy, T., Lieb, E.H. and Shastry, S.: The XY model has long-range order for all spins and all dimensions greater than one. Phys. Rev. Lett. **61**, 2582–2585 (1988)
- [L] Lieb, E.H.: The Bose fluid. In: *Lecture Notes in Theoretical Physics VIIC*, edited by W.E. Brittin. Univ. of Colorado Press, 1964, pp. 175–224
- [LS] Lieb, E.H. and Sakakura, A.Y.: Simplified approach to the ground state energy of an imperfect Bose gas II. Charged Bose gas at high density. Phys. Rev. A **133**, 899–906 (1964)
- [LLe] Lieb, E.H. and Lebowitz, J.L.: The constitution of matter: Existence of thermodynamics for systems composed of electrons and nuclei. Adv. in Math. **9**, 316–398 (1972)
- [LLi] Lieb, E.H. and Liniger, W.: Exact analysis of an interacting Bose gas I. The general solution and the ground state. Phys. Rev. **130**, 1605–1616 (1963). See Fig. 3
- [LN] Lieb, E.H. and Narnhofer, H.: The thermodynamic limit for jellium. J. Stat. Phys. **12**, 295–310 (1975); Errata. **14**, 465 (1976)
- [LY] Lieb, E.H. and Yngvason, J.: Ground state energy of the low density Bose gas. Phys. Rev. Lett. **80**, 2504–2507 (1998)

Communicated by M. Aizenman

Correction to “Ground State Energy of the One-Component Charged Bose Gas”

Elliott H. Lieb and Jan Philip Solovej

The proof of Lemma B.1 of [1] contains an unjustified operator inequality. In the last estimate on p. 162 the Cauchy-Schwarz inequality was used incorrectly. The lemma is however still correct as stated. We shall show this below. The operator inequality to be proven is that

$$(\nabla\theta)^2 \frac{-\Delta_N}{-\Delta_N + s^{-2}} + \frac{-\Delta_N}{-\Delta_N + s^{-2}} (\nabla\theta)^2 \leq Ct^{-2} \frac{-\Delta_N}{-\Delta_N + s^{-2}} + Cs^2 t^{-4}, \quad (1)$$

where $-\Delta_N$ is the Neumann Laplacian of some bounded open set $\mathcal{O} \subset \mathbb{R}^n$, $s > 0$, and $\theta \in C^\infty(\overline{\mathcal{O}})$ is constant near the boundary of $\mathcal{O}$ and satisfies the estimates $\|\partial^\alpha \theta\|_\infty \leq Ct^{-|\alpha|}$, for some $t > 0$ and all multi-indices α with $|\alpha| \leq 3$.

The proof of (1) is a little technical. For the application in the paper the following estimate, in which the Cauchy-Schwarz inequality has been used correctly, would have sufficed.

$$\begin{aligned} (\nabla\theta)^2 \frac{-\Delta_N}{-\Delta_N + s^{-2}} + \frac{-\Delta_N}{-\Delta_N + s^{-2}} (\nabla\theta)^2 &\leq st(\nabla\theta)^4 + (st)^{-1} \left(\frac{-\Delta_N}{-\Delta_N + s^{-2}} \right)^2 \\ &\leq Cst^{-3} + C(st)^{-1} \frac{-\Delta_N}{-\Delta_N + s^{-2}}. \end{aligned}$$

In order to prove (1) we shall use the two operator inequalities

$$[-\Delta_N, f][f, -\Delta_N] \leq C\|\nabla f\|_\infty^2 (-\Delta_N) + C \left(\sum_i \|\partial_i^2 f\|_\infty \right)^2 \quad (2)$$

and

$$\begin{aligned} f(-\Delta_N)f &= -\sum_i \partial_i f^2 \partial_i + \sum_i [\partial_i f, f \partial_i] \leq -C \sum_i \partial_i f^2 \partial_i + C \sum_i (\partial_i f)^2 \\ &\leq C\|f\|_\infty^2 (-\Delta_N) + C\|\nabla f\|^2, \end{aligned} \quad (3)$$

where f is a smooth function with compact support in $\mathcal{O}$, which we identify as a multiplication operator.

We begin by rewriting the left side of (1).

$$\begin{aligned} &(\nabla\theta)^2 \frac{-\Delta_N}{-\Delta_N + s^{-2}} + \frac{-\Delta_N}{-\Delta_N + s^{-2}} (\nabla\theta)^2 \\ &= \int_0^\infty (\nabla\theta)^2 \frac{-\Delta_N}{(-\Delta_N + s^{-2} + u)^2} + \frac{-\Delta_N}{(-\Delta_N + s^{-2} + u)^2} (\nabla\theta)^2 du \\ &= \int_0^\infty \left(\frac{1}{-\Delta_N + s^{-2} + u} (\nabla\theta)^2 \frac{-\Delta_N}{-\Delta_N + s^{-2} + u} + \frac{-\Delta_N}{-\Delta_N + s^{-2} + u} (\nabla\theta)^2 \frac{1}{-\Delta_N + s^{-2} + u} \right) du \\ &\quad + \int_0^\infty \left(\frac{1}{-\Delta_N + s^{-2} + u} [-\Delta_N, (\nabla\theta)^2] \frac{-\Delta_N}{(-\Delta_N + s^{-2} + u)^2} \right. \\ &\quad \left. + \frac{-\Delta_N}{(-\Delta_N + s^{-2} + u)^2} [(\nabla\theta)^2, -\Delta_N] \frac{1}{-\Delta_N + s^{-2} + u} \right) du. \end{aligned} \quad (4)$$

The first integral we estimate using a Cauchy-Schwarz inequality

$$\begin{aligned}
 & \int_0^\infty \frac{1}{-\Delta_N + s^{-2} + u} (\nabla\theta)^2 \frac{-\Delta_N}{-\Delta_N + s^{-2} + u} + \frac{-\Delta_N}{-\Delta_N + s^{-2} + u} (\nabla\theta)^2 \frac{1}{-\Delta_N + s^{-2} + u} du \\
 & \leq t^2 \int_0^\infty \frac{1}{-\Delta_N + s^{-2} + u} (\nabla\theta)^2 (-\Delta_N) (\nabla\theta)^2 \frac{1}{-\Delta_N + s^{-2} + u} + t^{-2} \frac{-\Delta_N}{(-\Delta_N + s^{-2} + u)^2} du \\
 & \leq Ct^{-2} \frac{-\Delta_N}{-\Delta_N + s^{-2}} + Cs^2 t^{-4},
 \end{aligned}$$

where in the last estimate we have used (3) with $f = (\nabla\theta)^2$.

The last integral in (4) we estimate again using a Cauchy-Schwarz inequality this time together with (2) with $f = (\nabla\theta)^2$

$$\begin{aligned}
 & \int_0^\infty \left(\frac{1}{-\Delta_N + s^{-2} + u} [-\Delta_N, (\nabla\theta)^2] \frac{-\Delta_N}{(-\Delta_N + s^{-2} + u)^2} \right. \\
 & \quad \left. \frac{-\Delta_N}{(-\Delta_N + s^{-2} + u)^2} [(\nabla\theta)^2, -\Delta_N] \frac{1}{-\Delta_N + s^{-2} + u} \right) du \\
 & \leq t^4 \int_0^\infty \frac{1}{-\Delta_N + s^{-2} + u} [-\Delta_N, (\nabla\theta)^2] [(\nabla\theta)^2, -\Delta_N] \frac{1}{-\Delta_N + s^{-2} + u} du \\
 & \quad + t^{-4} \int_0^\infty \frac{(-\Delta_N)^2}{(-\Delta_N + s^{-2} + u)^4} du \\
 & \leq t^4 \int_0^\infty \frac{1}{-\Delta_N + s^{-2} + u} (Ct^{-6}(-\Delta_N) + Ct^{-8}) \frac{1}{-\Delta_N + s^{-2} + u} du + s^2 t^{-4} \\
 & \leq Ct^{-2} \frac{-\Delta_N}{-\Delta_N + s^{-2}} + Cs^2 t^{-4}.
 \end{aligned}$$

This proves (1).

References

- [1] Elliott H. Lieb and Jan Philip Solovej, Ground State Energy of the One-Component Charged Bose Gas, *Commun. Math. Phys.* **217**, 127–163 (2001).

Ground State Energy of the Two-Component Charged Bose Gas*

Elliott H. Lieb^{1, **}, Jan Philip Solovej^{2, ***}, †

¹ Departments of Physics and Mathematics, Jadwin Hall, Princeton University, PO Box 708, Princeton, NJ 08544-0708, USA. E-mail: lieb@princeton.edu

² School of Mathematics, Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA. E-mail: solovej@math.ku.dk

Received: 8 April 2004 / Accepted: 10 June 2004

Published online: 8 July 2004 – © E.H. Lieb and J.P. Solovej 2003

Dedicated to Freeman J. Dyson on the occasion of his 80th birthday

Abstract: We continue the study of the two-component charged Bose gas initiated by Dyson in 1967. He showed that the ground state energy for N particles is at least as negative as $-CN^{7/5}$ for large N and this power law was verified by a lower bound found by Conlon, Lieb and Yau in 1988. Dyson conjectured that the exact constant C was given by a mean-field minimization problem that used, as input, Foldy's calculation (using Bogolubov's 1947 formalism) for the one-component gas. Earlier we showed that Foldy's calculation is exact insofar as a lower bound of his form was obtained. In this paper we do the same thing for Dyson's conjecture. The two-component case is considerably more difficult because the gas is very non-homogeneous in its ground state.

1. Introduction

In 1967 Dyson [D] showed that a system composed of non-relativistic, charged bosons is unstable in the sense that the ground state energy of N particles is at least as negative as $-CN^{7/5}$ instead of $-CN$, where C is some constant. A lower bound of the form $-C'N^{7/5}$ was derived later [CLY], thereby establishing the correctness of the exponent $7/5$, but not the constant C .

In an earlier, parallel development, in 1961 Foldy [F] considered the problem of the one-component Bose gas ("jellium") in which charged particles (all of the same charge)

* © 2003 by the authors. This article may be reproduced in its entirety for non-commercial purposes.

** Work partially supported by U.S. National Science Foundation grant PHY01 39984-A01.

*** Work partially supported by NSF grant DMS-0111298, by EU grant HPRN-CT-2002-00277, by MaPhySto – A Network in Mathematical Physics and Stochastics, funded by The Danish National Research Foundation, and by grants from the Danish research council.

† On leave from Department of Mathematics, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark.

move in a uniformly charged, neutralizing background. Using Bogolubov’s 1947 theory [B], Foldy “derived” the high density asymptotics for the ground state energy of this problem as proportional to $-I_0 N \rho^{1/4}$ where I_0 is defined in Eq. (5) below. The correctness of this $\rho^{1/4}$ law, but not the coefficient I_0 , was also proved in ([CLY]).

Dyson was motivated by Foldy’s work, for he realized that if one treated one of the two components (say, the positive one) as a background for the other, and if one allowed the density to be variable, one would easily arrive — heuristically, at least — at the $N^{7/5}$ law. This will be explained below.

Two obvious questions arise from this earlier work. What is the correct coefficient for the $\rho^{1/4}$ law at high density and what is the correct coefficient for the $N^{7/5}$ law? The former question was resolved by us in [LSol], where we showed that Foldy’s I_0 is, indeed, correct at high density as a lower bound. Foldy’s calculation is “essentially” an upper bound, but some technical issues must be clarified. The proof that Foldy’s calculation, indeed, gives an upper bound can be found in [S].

In [D] Dyson derives a rigorous upper bound for the $N^{7/5}$ law, but with a coefficient C that is clearly too small. He conjectures a “correct” coefficient, however, and in the present paper we shall show that Dyson’s conjectured coefficient gives a correct lower bound (asymptotically as $N \rightarrow \infty$). An asymptotically correct upper bound for the two-component gas energy is also given in [S].

Actually, our lower bound is slightly more general than just the case of $N/2$ particles of each charge. We prove the lower bound for the case in which the total number is N without restriction to the $N/2$ case.

In order to understand the reason that the proof for the two-component case is more difficult than that for jellium, presented in [LSol], it is necessary to recapitulate Dyson’s argument briefly. His picture is that there is a local density of particles $\rho(x)$, which has a local energy density given by Foldy’s formula, i.e., $I_0 \rho(x)^{5/4}$. One might question whether the jellium energy can be simply taken over to the two-component situation, but it is correct to do so, as our lower bound shows. There is a good reason for this within the Bogolubov theory, as we shall explain in Sect. 3, but let us continue with Dyson’s picture now. In addition to the local energy there is also a kinetic energy caused by the variation in ρ , namely $\int |\nabla \sqrt{\rho(x)}|^2 dx$. Such an “envelope” energy is familiar from Thomas-Fermi-Weizsäcker and Gross-Pitaevskii theories, for example.

If this total energy is minimized with respect to ρ we are led to a differential equation in $\sqrt{\rho}$ with the side condition that $\int \rho = N$, but the basic features are clear. The scale length of $\rho(x)$ will be of the order $N^{-1/5}$, the amplitude of ρ will be $N^{8/5}$ and the energy will be $N^{7/5}$. Indeed, if we define $\Phi(x) = N^{-4/5} \sqrt{\rho(N^{-1/5}x)}$ such that $\int \Phi^2 = 1$ then the energy as a function of ρ that we have to minimize can be rewritten as

$$\frac{1}{2} \int_{\mathbb{R}^3} |\nabla \sqrt{\rho}|^2 - I_0 \int_{\mathbb{R}^3} \rho^{5/4} = N^{7/5} \left(\frac{1}{2} \int_{\mathbb{R}^3} |\nabla \Phi|^2 - I_0 \int_{\mathbb{R}^3} \Phi^{5/2} \right), \tag{1}$$

which makes the scaling explicit. We then have to minimize the right side with respect to Φ under the condition $\int \Phi^2 = 1$. The finiteness of this minimum energy is an easy consequence of the Sobolev inequality. The existence of a minimizing Φ is a little harder and follows with the help of rearrangement inequalities. It satisfies a Lane-Emden differential equation for some $\mu > 0$ and all $x \in \mathbb{R}^3$,

$$-\Delta \Phi(x) - \frac{5}{2} I_0 \Phi(x)^{3/2} + \mu \Phi(x) = 0. \tag{2}$$

The uniqueness of Φ , up to translations, is harder still. See [Be, K, MS, Z].

Dyson's conjecture is contained in [D, Eqs. 104, 105]. The normalization convention and units employed there are not completely clear, but the heuristics leading to (104) is clear. This will be clarified in Sect. 3.

The last topic to discuss in this introduction is the essential difficulty inherent in the two-component problem. As in the jellium proof in [LSo] we decompose space into suitable small boxes and impose Neumann boundary conditions on each. The scale length of these boxes is $N^{-2/5+\varepsilon}$. Using the "sliding argument" of [CLY] we can ignore the Coulomb interaction between different boxes. We must then distribute the N particles in the boxes and what we must prove is that the number in the various boxes has a coarse grained density given by the solution ρ to (1). Since the energy is super concave (i.e., $-N^{7/5}$), the lowest energy is obtained by putting all the N particles in one box. What prevents this from happening is that the boxes are not really totally independent, thanks to the kinetic energy operator. In other words, we must somehow save a little bit of the kinetic energy operator to prevent wild variations in particle density between neighboring boxes.

The conundrum is that the mean-field energy in (1) uses all the kinetic energy, not just some of it. Likewise, to get the intrabox energy (the second term in (1)) we also need the full kinetic energy. The resolution is to split the kinetic energy operator $-\Delta$ into a high-momentum part for use in calculating the intrabox energy and a low-momentum part for use in reproducing the first term in (1). Naturally, error terms will arise and the chief difference between this paper and our earlier jellium paper is centered on the definition of the splitting and the management of the induced error terms.

The proof of the main theorem starts in Sect. 4 where we show how to localize the problem into large boxes of size $L \gg N^{-1/5}$ with Dirichlet boundary conditions. This size is larger than the expected size of the bound complex. The decomposition of the kinetic energy into large and small momentum is carried out in an appendix. It is used in Sect. 5, which localizes further into really small boxes of size $\ell \gg N^{-2/5}$. For the relevance of the scale $N^{-2/5}$ the reader is referred to the heuristic discussion in Sect. 3. The control of the electrostatics using sliding is also discussed in both of these sections. Section 6 discusses the ultraviolet and infrared cutoffs of the interaction potential with control of the errors they introduce. In contrast to the treatment in [LSo] we now need to use the kinetic energy to control the errors caused by the ultraviolet and infrared cutoffs in the potential. Section 7 controls all the unimportant parts of the localized Hamiltonian and reduces the problem to Bogolubov's Hamiltonian, which is analyzed in Sect. 8.

Section 9 gives the first, simple bound on kinetic energy, local non-neutrality, and an estimate on the local condensation. Section 10 improves the estimate on condensation with the help of the method of "localizing large matrices" in [LSo] (and which is reviewed in an appendix). (Recall that in [LSo] we had to reach the final estimate on the various energies by a succession of finer and finer error bounds, each taking the previous bound as input.) In Sect. 11, we give the final bound on the energy in each small box. We have to treat boxes with few particles as well as many particles separately. In Sect. 12 we show how the kinetic energy estimate in the appendix (in which the low momentum kinetic energy *between* boxes leads to a difference energy on a lattice) leads, in turn, to the term $\int |\nabla \sqrt{\rho}|^2$ in the energy functional (1). In the final Sect. 13 all the estimates are put together and we show how to choose the various parameters to get the desired minimization problem for the lower bound.

We thank Y.Y. Li and M.I. Weinstein for pointing out the references [K, MS, Z] to us.

2. Basic Definitions and Main Theorem

We consider N particles with charges $e_i = \pm 1, i = 1, \dots, N$. The Hamiltonian describing the system is

$$H_N = \sum_{i=1}^N -\frac{1}{2}\Delta_i + \sum_{1 \leq i < j \leq N} \frac{e_i e_j}{|x_i - x_j|} \quad (3)$$

acting on $L^2(\mathbb{R}^{3N})$. We shall not specify the number of positive or negative particles, but simply consider the smallest possible energy

$$E(N) = \inf\{\inf \text{spec}_{L^2(\mathbb{R}^{3N})} H_N \mid e_i = \pm 1, i = 1, \dots, N\}.$$

Instead of considering H_N depending on the parameters $e_1, \dots, e_N$ we may consider it as *one* operator on the enlarged space $L^2((\mathbb{R}^3 \times \{1, -1\})^N)$, where the set $\{1, -1\}$ contains the values of the charge variables. Then

$$E(N) = \inf \text{spec}_{L^2((\mathbb{R}^3 \times \{1, -1\})^N)} H_N.$$

Thus $E(N)$ is the infimum of $\langle \Psi, H_N \Psi \rangle$, over normalized functions Ψ in $L^2((\mathbb{R}^3 \times \{1, -1\})^N)$. Here we may restrict Ψ to be non-negative and thus moreover also to be symmetric under the interchange of particles. Thus $E(N)$ is the energy of a charged Bose gas.

Our main result in this paper is the asymptotic *lower* bound on $E(N)$ conjectured by Dyson in [D]. The corresponding asymptotic *upper* bound is given in [S]. Together these results prove Dyson's formula.

Theorem 2.1 (Dyson's formula). *As $N \rightarrow \infty$ we have*

$$E(N) = -AN^{7/5} + o(N^{7/5}),$$

where A is the positive constant determined by the variational principle

$$-A = \inf \left\{ \frac{1}{2} \int_{\mathbb{R}^3} |\nabla \Phi|^2 - I_0 \int_{\mathbb{R}^3} \Phi^{5/2} \mid 0 \leq \Phi, \int_{\mathbb{R}^3} \Phi^2 \leq 1 \right\} \quad (4)$$

with

$$I_0 = (2/\pi)^{3/4} \int_0^\infty \left(1 + x^4 - x^2 (x^4 + 2)^{1/2} \right) dx = \frac{2^{3/2} \Gamma(3/4)}{5\pi^{1/4} \Gamma(5/4)}. \quad (5)$$

It will be clear from the proof (see the discussion in Sect. 13) that the error in Dyson's formula could have been written in the form $N^{7/5-\xi}$ for some $\xi > 0$. Although this is in principle straightforward we have not attempted to optimize the error term to determine the exact exponent ξ .

3. Heuristic Derivation of the Energy

In this section we give the heuristic derivation that leads to the local energy density $I_0 \rho^{5/4}$ in the second term in (1) (see also (4)). The fact that the constant for the two-component gas is the same as I_0 , the constant for the one-component gas, is somewhat mysterious. After all, there is not only the negative $+-$ interaction energy, but there are also the $++$ and $--$ energies. Moreover, the ρ in the two-component case refers to the total number of particles, which is twice the number of each charge (in the neutral case), whereas ρ refers only to the movable particles of one sign in the one-component case.

Another problem is that the usual thermodynamic limit exists for the one-component gas only because we constrain the background to have uniform density and we do not allow it to contract to a high density ball, as it would if one merely minimizes the energy. The two-component gas does not have the usual thermodynamic limit because, as Dyson showed, its energy goes as $-N^{7/5}$.

Nevertheless, we shall describe the two-component charged gas on a small local scale ℓ as a gas of uniform density ρ . For such a uniform gas we shall imitate Foldy's calculation [F] to arrive at the energy $-I_0 \rho^{5/4} \ell^3$. We shall assume that $\ell \ll N^{-1/5}$ (the scale on which $\rho(x)$ varies), but $\ell \gg \rho^{-1/4}$ which, as we shall see, is the relevant scale for the uniform gas. Note that we expect $\rho \sim N^{8/5}$ and hence $\rho^{-1/4} \sim N^{-2/5}$. In any event, we are being "heuristic" in this section and the reader is welcome to ignore this "derivation" if it is not pleasing.

We shall use periodic boundary conditions, as usual, and write the Hamiltonian in second quantized form in the manner of Bogolubov/Foldy as

$$\begin{aligned}
 H_F = & \sum_k \frac{1}{2} k^2 (a_{k+}^* a_{k+} + a_{k-}^* a_{k-}) \\
 & + \frac{1}{2\ell^3} \sum_{k \neq 0} \sum_{pq} \frac{4\pi}{k^2} \left(a_{p+}^* a_{q+}^* a_{(q-k)+} a_{(p+k)+} + a_{p-}^* a_{q-}^* a_{(q-k)-} a_{(p+k)-} \right. \\
 & \left. - 2a_{p+}^* a_{q-}^* a_{(q-k)-} a_{(p+k)+} \right), \tag{6}
 \end{aligned}$$

where the sums are over momenta in the set $\frac{2\pi}{\ell} \mathbb{Z}^3$. Here $a_{p\pm}^*$ creates a state with momentum p and charge ± 1 .

In the second term we have excluded $k = 0$, which is also what Foldy does. It would of course be meaningless to include this term and leave the Fourier transform of the potential equal to $\frac{4\pi}{k^2}$. If, instead, we defined the Fourier transform of the potential for $k = 0$ to be proportional to ℓ^2 (which is the maximal value of the Fourier transform for $k \neq 0$), the contribution from the term $k = 0$ would be proportional to $-(\rho \ell^3) \ell^{-1}$ for a neutral system. If, as we assume, $\ell \gg \rho^{-1/4}$ then $(\rho \ell^3) \ell^{-1} \ll \rho^{5/4} \ell^3$. Hence we may ignore the $k = 0$ contribution. In Foldy's situation the corresponding term would not contribute to the thermodynamic limit.

The next step in the heuristic derivation is to exclude those terms in the second sum above that do not contain precisely two creation or annihilation operators of particles of momentum zero. Subsequently these zero momentum creation or annihilation operators $a_{0\pm}^*$ and $a_{0\pm}$ are replaced by the square root of half the particle number, namely $\sqrt{\rho \ell^3}/2$. We then arrive at

$$\begin{aligned} & \sum_k \frac{1}{4} k^2 (a_{k+}^* a_{k+} + a_{-k+}^* a_{-k+} + a_{k-}^* a_{k-} + a_{-k-}^* a_{-k-}) \\ & + \frac{\rho}{4} \sum_{k \neq 0} \frac{4\pi}{k^2} \left[a_{k+}^* a_{k+} + a_{-k+}^* a_{-k+} + a_{k+}^* a_{-k+}^* + a_{k+} a_{-k+} \right. \\ & \quad + a_{k-}^* a_{k-} + a_{-k-}^* a_{-k-} + a_{k-}^* a_{-k-}^* + a_{k-} a_{-k-} \\ & \quad - (a_{k+}^* a_{k-} + a_{-k+}^* a_{-k-} + a_{k-}^* a_{k+} + a_{-k-}^* a_{-k+} + a_{k+}^* a_{-k-}^* \\ & \quad \left. + a_{-k+}^* a_{k-}^* + a_{k+} a_{-k-} + a_{-k+} a_{k-}) \right]. \end{aligned}$$

This expression should be compared with Lemma 8.1, where we arrive at a similar expression as a rigorous lower bound on part of the full Hamiltonian. (In comparing with the lemma one should replace $b_k \rightarrow a_k$, $v^+, v^- \rightarrow \rho \ell^3/2$, $\hat{V}_{r,R}(k) \rightarrow \frac{4\pi}{|k|^2}$, $\gamma_{\varepsilon,t} \rightarrow 1$, $(\ell t^6)^{-2} \rightarrow 0$, and of course also $\frac{\ell^3}{(2\pi)^3} \int dk \rightarrow \sum_k$.)

The final step is to recognize that the resulting quadratic Hamiltonian has the following property, as the reader can easily check: The operators $a_{k\pm}$ always appear in the potential energy term (the last sum above) in the combination $d_k = (a_{k+} - a_{k-})/\sqrt{2}$. This is a normal mode since $[d_k, d_q^*] = \delta_{k,q}$. The other normal mode $c_k = (a_{k+} + a_{k-})/\sqrt{2}$ appears only in the kinetic energy term (the first sum above), i.e., the kinetic energy is $a_{k+}^* a_{k+} + a_{k-}^* a_{k-} = c_k^* c_k + d_k^* d_k$. The ground state is achieved by having no c_k excitations, which leaves us just with the term $d_k^* d_k$ in the kinetic energy term.

The conclusion is that the quadratic Hamiltonian is now exactly the same as Foldy's (but with d_k in place of a_k) and, therefore, the ground state energy is $-I_0 \rho^{5/4} \ell^3$. This conclusion could also have been arrived at by an explicit diagonalization of the total quadratic Hamiltonian. (See Sect. 8, in particular Theorem 8.2, for comparison.)

The detailed diagonalization analysis shows that the relevant momenta k are of magnitude $\rho^{1/4}$ and hence as mentioned above the relevant length scale is $\rho^{-1/4}$. The assumption $\ell \gg \rho^{-1/4}$ allows one to replace sums over the lattice $\frac{2\pi}{\ell} \mathbb{Z}^3$ by integrals $\frac{\ell^3}{(2\pi)^3} \int_{\mathbb{R}^3} dk$. This is how the integral in (5) appears.

4. Localization

If Dyson's conjecture is correct then the size of the boson cloud is proportional to $N^{-1/5}$. As a first step we shall localize the problem into cubes of size L , where we choose L as a function of N in such a way that $N^{1/5} L \rightarrow \infty$ as $N \rightarrow \infty$. Exactly how $N^{1/5} L \rightarrow \infty$ will be determined at the end of the analysis in Sect. 13. As a consequence of our results we shall see that essentially all particles will concentrate within one of these cubes of size L .

We shall do the localization in such a way that the cubes do not interact and the analysis can be done in each cube independently. The only thing to bear in mind is that the total number of particles in all cubes is N .

In analyzing the individual cubes we shall perform a further localization into smaller cells of a size $\ell < L$ depending on N in such a way that $N^{2/5} \ell \rightarrow \infty$ as $N \rightarrow \infty$ (precisely how will again be determined in Sect. 13).

Ground State Energy of the Two-Component Charged Bose Gas

We shall first describe how we control the electrostatic interaction between the different regions into which we localize. We do this in a manner very similar to what was done in [LS0] using the sliding technique of [CLY], and shall use this technique both for the localization into the large cubes and again when we localize into the smaller cells.

Let t , with $0 < t < 1/2$, be a parameter which will be chosen later in Sect. 13 to depend on N in such a way that $t \rightarrow 0$ as $N \rightarrow \infty$.

Let $\Theta, \theta \in C_0^\infty(\mathbb{R}^3)$ satisfy

1. $0 \leq \Theta, \theta \leq 1$, $\theta(x) = \theta(-x)$, and $\Theta(x) = \Theta(-x)$.
2. $\text{supp } \theta \subset [(-1+t)/2, (1-t)/2]^3$, $\text{supp } \Theta \subset [(-1-t)/2, (1+t)/2]^3$.
3. $\theta(x) = 1$ for $x \in [(-1+2t)/2, (1-2t)/2]^3$, and $\Theta(x) = 1$ for $x \in [(-1+t)/2, (1-t)/2]^3$.
4. All derivatives of order m for $m \leq 3$ of the functions $\theta, \sqrt{1-\theta^2}, \Theta$ are uniformly bounded by Ct^{-m} , where C is some universal constant.
5. For all $x \in \mathbb{R}^3$ we have $\sum_{k \in \mathbb{Z}^3} \Theta(x-k)^2 = 1$.

We introduce the two constants $\gamma, \tilde{\gamma}$ such that $\gamma \int \theta(y)^2 dy = 1$ and $\tilde{\gamma} \int \Theta^4(y)^2 dy = 1$. Then

$$1 \leq \gamma \leq (1-2t)^{-3}, \quad (1+t)^{-3} \leq \tilde{\gamma} \leq (1-t)^{-3}. \quad (7)$$

We also introduce the Yukawa potential $Y_m(x) = |x|^{-1} e^{-m|x|}$ for $m \geq 0$. For $m = 0$ this is, of course, the Coulomb potential.

Lemma 4.1 (Electrostatic decoupling of boxes using sliding). *There exists a function of the form $\omega(t) = Ct^{-4}$ (we assume that $\omega(t) \geq 1$ for $t < 1/2$) such that for all $x_1, x_2, \dots, x_N \in \mathbb{R}^3$, all $e_1, e_2, \dots, e_N$, with $|e_i| = 1$ for $i = 1, 2, \dots, N$, all $m \geq 0$, and all $\lambda > 0$ we have*

$$\begin{aligned} & \sum_{1 \leq i < j \leq N} e_i e_j Y_m(x_i - x_j) \\ & \geq \gamma \int_{\mathbb{R}^3} \sum_{1 \leq i < j \leq N} e_i e_j \theta\left(\frac{x_i}{\lambda} - z\right) Y_{m+\frac{\omega(t)}{\lambda}}(x_i - x_j) \theta\left(\frac{x_j}{\lambda} - z\right) dz - \frac{N\omega(t)}{2\lambda} \end{aligned} \quad (8)$$

and likewise

$$\begin{aligned} & \sum_{1 \leq i < j \leq N} e_i e_j Y_m(x_i - x_j) \\ & \geq \tilde{\gamma} \int_{\mathbb{R}^3} \sum_{1 \leq i < j \leq N} e_i e_j \Theta^2\left(\frac{x_i}{\lambda} - z\right) Y_{m+\frac{\omega(t)}{\lambda}}(x_i - x_j) \Theta^2\left(\frac{x_j}{\lambda} - z\right) dz - \frac{N\omega(t)}{2\lambda}. \end{aligned} \quad (9)$$

Proof. Since θ and Θ^2 have the same properties it is enough to consider θ . By rescaling we may assume that $\lambda = 1$. We have that

$$\int \gamma \theta(x-z) Y_{m+\omega}(x-y) \theta(y-z) dz = h(x-y) Y_{m+\omega}(x-y),$$

where we have set $h = \gamma \theta * \theta$. We chose γ such that $1 = h(0) = \gamma \int \theta(y)^2 dy$. Then h satisfies all the assumptions in Lemma 2.1 in [CLY]. We then conclude from Lemma 2.1

in [CLY] that the Fourier transform of the function $F(x) = Y_m(x) - h(x)Y_{m+\omega}(x)$ is non-negative, if ω is chosen large enough depending on h . [The detailed bounds from [CLY] show that we may in fact choose $\omega = Ct^{-4}$, since ω has to control the 4th derivative of h .] Note, moreover, that $\lim_{x \rightarrow 0} F(x) = \omega$. Hence

$$\sum_{1 \leq i < j \leq N} e_i e_j F(y_i - y_j) \geq - \sum_{i=1}^N \frac{e_i^2 \omega}{2} = - \frac{N\omega}{2}.$$

The lemma follows by writing $Y_m(x) = F(x) + h(x)Y_{m+\omega}(x)$. $\square$

We shall use the electrostatic decoupling (9) when we localize into the large cubes of size L and then (8) when we localize further into the smaller cells. We begin with the localization into large cubes.

Theorem 4.2 (Localization into a large cube). *Let*

$$H_{N,L} = \sum_{i=1}^N -\frac{1}{2} \Delta_{i,D} + \sum_{1 \leq i < j \leq N} e_i e_j \tilde{\gamma} Y_{\frac{\omega(t)}{L}}(x_i - x_j)$$

be a Hamiltonian acting in the space $L^2([-L/2, L/2]^3 \times \{1, -1\})^N$. Here Δ_D refers to the Dirichlet Laplacian in the cube $[-L/2, L/2]^3$. Let

$$E_L(N) = \inf \text{spec} H_{N,L}. \quad (10)$$

Then

$$E(N) \geq E_L(N) - N^{7/5} \left(Ct^{-2} (N^{1/5} L)^{-2} + \frac{1}{2} \omega(t) N^{-1/5} (N^{1/5} L)^{-1} \right).$$

Proof. For all $z \in \mathbb{R}^3$ let $\Theta_z(x) = \Theta((x/L) - z)$. We consider Θ_z as a multiplication operator on $L^2(\mathbb{R}^3)$. We have for all $z \in \mathbb{R}^3$ that

$$\sum_{q \in \mathbb{Z}^3} \Theta_{z+q}^2(x) = 1.$$

Using (9) with $m = 0$ and that $\int_{\mathbb{R}^3} f(z) dz = \sum_{q \in \mathbb{Z}^3} \int_{[-1/2, 1/2]^3} f(z+q) dz$ we see that

$$\begin{aligned} \sum_{1 \leq i < j \leq N} \frac{e_i e_j}{|x_i - x_j|} &\geq \sum_{\mathbf{k}=(k_1, \dots, k_N) \in \mathbb{Z}^{3N}} \int_{[-1/2, 1/2]^3} F_{\mathbf{k},z}(x_1, \dots, x_N)^2 \\ &\times \sum_{1 \leq i < j \leq N} \delta_{k_i k_j} e_i e_j \tilde{\gamma} Y_{\omega(t)/L}(x_i - x_j) dz - \frac{N\omega(t)}{2L}, \end{aligned}$$

where

$$F_{\mathbf{k},z}(x_1, \dots, x_N) = \Theta_{k_1+z}(x_1) \cdots \Theta_{k_N+z}(x_N), \quad \mathbf{k} = (k_1, \dots, k_N) \in \mathbb{Z}^{3N}.$$

We localize the kinetic energy using the formula

$$-\Delta = \sum_{q \in \mathbb{Z}^3} \Theta_{z+q}(-\Delta) \Theta_{z+q} - \sum_{q \in \mathbb{Z}^3} (\nabla \Theta_{z+q})^2 \geq \sum_{q \in \mathbb{Z}^3} \Theta_{z+q}(-\Delta) \Theta_{z+q} - C(tL)^{-2},$$

Ground State Energy of the Two-Component Charged Bose Gas

which gives

$$\begin{aligned} \sum_{i=1}^N -\Delta_i &\geq \sum_{\mathbf{k} \in \mathbb{Z}^{3N}} \int_{[-1/2, 1/2]^3} F_{\mathbf{k}, z}(x_1, \dots, x_N) \\ &\times \sum_{i=1}^N (-\Delta_i) F_{\mathbf{k}, z}(x_1, \dots, x_N) dz - CN(tL)^{-2}. \end{aligned}$$

The previous estimates imply that

$$\begin{aligned} H_N &\geq \sum_{\mathbf{k} \in \mathbb{Z}^{3N}} \int_{[-1/2, 1/2]^3} F_{\mathbf{k}, z}(x_1, \dots, x_N) \\ &\times \sum_{q \in \mathbb{Z}^3} H_{\mathbf{k}, z}^{(q)} F_{\mathbf{k}, z}(x_1, \dots, x_N) dz - \frac{N\omega(t)}{2L} - CN(tL)^{-2}, \end{aligned}$$

where we have introduced the operator

$$H_{\mathbf{k}, z}^{(q)} = \sum_{i=1}^N -\frac{1}{2} \delta_{k_i, q} \Delta_{i, D}^{(q+z)} + \sum_{1 \leq i < j \leq N} \delta_{k_i, q} \delta_{k_j, q} e_i e_j \tilde{\gamma} Y_{\omega(t)/L}(x_i - x_j),$$

where $\Delta_D^{(q+z)}$ denotes the Dirichlet Laplacian in the cube $\{(q+z)L\} + [-L/2, L/2]^3$. Note that the operator $H_{\mathbf{k}, z}^{(q)}$ above acts on functions for which the space variables are in the set $\{(q+z)L\} + [-L/2, L/2]^3$. The operator $H_{\mathbf{k}, z}^{(q)}$ is hence unitary equivalent to $H_{N, L}$ with N replaced by

$$N_q(\mathbf{k}) = \#\mathcal{N}_q(\mathbf{k}), \quad \mathcal{N}_q(k_1, \dots, k_N) = \{i \mid k_i = q\}.$$

We therefore get the lower bound

$$H_N \geq \sum_{\mathbf{k} \in \mathbb{Z}^{3N}} \int_{[-1/2, 1/2]^3} F_{\mathbf{k}, z}(x_1, \dots, x_N)^2 \sum_{q \in \mathbb{Z}^3} E_L(N_q(\mathbf{k})) dz - \frac{N\omega(t)}{2L} - CN(tL)^{-2}.$$

The theorem follows if we can prove that for fixed $\mathbf{k} \in \mathbb{Z}^{3N}$,

$$\sum_{q \in \mathbb{Z}^3} E_L(N_q(\mathbf{k})) \geq E_L(N). \quad (11)$$

To conclude (11), let ψ_q be a normalized N_q -particle ground state eigenfunction of $H_{N_q, L}$, such that

$$(\psi_q, H_{N_q, L} \psi_q) = E_L(N_q).$$

(A normalized 0-particle function we consider simply to be the number 1.) Define the N -particle eigenfunction

$$\Psi_\tau(x_1, e_1, \dots, x_N, e_N) = \prod_q \psi_q((x_i, \tau_q e_i)_{i \in \mathcal{N}_q(\mathbf{k})}),$$

(for fixed $\mathbf{k}$ there are only finitely many factors different from 1 in the above product) where we shall choose the (finite) parameters $\tau_q \in \{1, -1\}$. We then find that

$$(\Psi_\tau, H_{N,L} \Psi_\tau) = \sum_{q \in \mathbb{Z}^3} E_L(N_q) + \frac{1}{2} \sum_{q \neq q'} \tau_q \tau_{q'} I_{qq'}, \tag{12}$$

(note that that there are again only finitely many non-zero terms in the sums) where

$$I_{qq'} = \left(\psi_q \psi_{q'}, \sum_i \sum_j \delta_{k_i,q} \delta_{k_j,q'} e_i e_j \widetilde{Y}_{\omega(t)/L}(x_i - x_j) \psi_q \psi_{q'} \right).$$

It is now easy to see that by an appropriate choice of the finite number of parameters τ_q , we can make sure that the last sum in (12) is non-positive. In fact, if we average over all possible choices of $\tau_q = \pm 1$ the last sum in (12) averages to 0.

The estimate (11) is now a consequence of the variational principle. $\square$

The rest of the analysis is concerned with estimating the energy $E_L(N)$ from (10). As explained in the beginning we may define $E_L(N)$ as $\inf \text{spec} H_{N,L}$ when we consider $H_{N,L}$ as an operator on the *symmetric* tensor product $\bigotimes_S^N L^2([-L/2, L/2]^3 \times \{1, -1\})$.

In the space $\bigotimes_S^N L^2(\mathbb{R}^3 \times \{1, -1\})$ we shall use the notation of second quantization. If $u \in L^2(\mathbb{R}^3 \times \{1, -1\})$ then $a^*(u)$ denotes the creation operator in the Fock space $\bigoplus_{N=0}^\infty \bigotimes_S^N L^2(\mathbb{R}^3 \times \{1, -1\})$. Products of the form $a_0^*(u) a_0(u')$ or $a_0(u') a_0^*(u)$ are however

bounded operators on each of the fixed particle number spaces $\bigotimes_S^N L^2(\mathbb{R}^3 \times \{1, -1\})$. We may consider $H_{N,L}$ as acting in this space but restricted to functions with support in $([-L/2, L/2]^3 \times \{1, -1\})^N$.

If $u \in L^2(\mathbb{R}^3)$ we use the notation $a_\pm^*(u)$ for the creation operator which creates the function u with charge ± 1 respectively, i.e., the function u defined on $\mathbb{R}^3 \times \{1\}$ or $\mathbb{R}^3 \times \{-1\}$ respectively.

5. The Localization of the Operator $H_{N,L}$

We turn to the second localization into smaller cubes of size $\ell > 0$, where we shall choose ℓ as a function of N in such a way that $\ell \leq L$ and $N^{2/5} \ell \rightarrow \infty$ as $N \rightarrow \infty$. As mentioned the precise choice of ℓ as a function of N will be made at the end of Sect. 13 when we complete the proof of Dyson’s formula.

Localizing the kinetic energy is difficult. In the paper [LSo] we used Neumann boundary conditions on the small cubes. In this way we got a lower bound by studying independent small boxes. In the present situation we will not get the correct answer if we bound the kinetic energy below by Neumann energies in independent boxes. As explained the “kinetic energy” between boxes is important. A lower bound on the kinetic energy which both contains a contribution from within each box and a contribution that connects boxes is given in Appendix 13. We shall use this result now.

We use the function θ to do the localization. For $z \in \mathbb{R}^3$ we define $\chi_z(x) = \theta((x/\ell) - z)$. Then $\text{supp } \chi_z \subset \{z\ell\} + [(-1+t)\ell/2, (1-t)\ell/2]^3$. Let $\eta_z = \sqrt{1 - \chi_z^2}$.

Ground State Energy of the Two-Component Charged Bose Gas

Then $\|\partial^\alpha \chi_z\|_\infty \leq C(\ell t)^{-|\alpha|}$ and $\|\partial^\alpha \eta_z\|_\infty \leq C(\ell t)^{-|\alpha|}$ for all multi-indices α with $|\alpha| \leq 3$.

Let X_z be the characteristic function of the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$, i.e., $X_z(x) = X_0(x - \ell z)$. Let $\mathcal{P}_z$ denote the projection onto the subspace of $L^2(\{z\ell\} + [-\ell/2, \ell/2]^3) \times \{1, -1\}$ consisting of functions orthogonal to constants. We shall consider $\mathcal{P}_z$ as a projection in $L^2(\mathbb{R}^3 \times \{1, -1\})$. We define for each $z \in \mathbb{R}^3$ the operator

$$\mathcal{K}^{(z)} = \mathcal{P}_z \chi_z \frac{(-\Delta)^2}{-\Delta + (\ell t^6)^{-2}} \chi_z \mathcal{P}_z. \quad (13)$$

The operators $\mathcal{K}^{(z)}$ shall play the role of the kinetic energy within each box.

Let $a_{0\pm}^*(z)$ be the creation operators

$$a_{0\pm}^*(z) = a_\pm^*(\ell^{-3/2} X_z), \quad (14)$$

i.e., the operator creating the constants in the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$. We introduce the notation

$$\hat{n}_z^\pm = a_{0\pm}^*(z) a_{0\pm}(z) \quad \text{and} \quad \hat{n}_z = \hat{n}_z^+ + \hat{n}_z^-. \quad (15)$$

We refer to $\hat{n}_z$ as the *number of particles in the condensate* in the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$. The operator for the *total number of particles* in the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$ is

$$\hat{v}_z = \sum_{i=1}^N X_z(x_i). \quad (16)$$

The operators $\hat{v}_z^\pm$ for the total numbers of positively and negatively charged particles are determined by

$$\hat{v}_z^+ + \hat{v}_z^- = \hat{v}_z \quad \text{and} \quad \hat{v}_z^+ - \hat{v}_z^- = \sum_{i=1}^N e_i X_z(x_i).$$

We refer to $\hat{v}_z - \hat{n}_z$ as the *number of excited particles*.

The kinetic energy that connects boxes will be a type of lattice Laplacian. In fact, for a map $S : \mathbb{Z}^3 \rightarrow \mathbb{R}$ we define a lattice Laplacian

$$T(S) = \sum_{\substack{\sigma_1, \sigma_2 \in \mathbb{Z}^3 \\ |\sigma_1 - \sigma_2| = \sqrt{2}}} \frac{1}{12} (S(\sigma_1) - S(\sigma_2))^2 + \sum_{\substack{\sigma_1, \sigma_2 \in \mathbb{Z}^3 \\ |\sigma_1 - \sigma_2| = \sqrt{3}}} \frac{1}{24} (S(\sigma_1) - S(\sigma_2))^2. \quad (17)$$

This specific form of T has been chosen because it is convenient when we shall later compare with a continuum Laplacian (see Sect. 12).

Using Theorem B.1 in Appendix 13 with $\Omega = [-(L + \ell)/(2\ell), (L + \ell)/(2\ell)]^3$ we obtain the following result.

Lemma 5.1 (Kinetic energy localization). *If $\Psi \in \bigotimes_S^N L^2(\mathbb{R}^3 \times \{1, -1\})$ is normalized and has support in $([-L/2, L/2]^3 \times \{1, -1\})^N$ then for all $\varepsilon > 0$, $0 < t < 1/2$,*

$$(1 + \varepsilon + Ct^3) \left(\Psi, \sum_{i=1}^N -\Delta_i \Psi \right) \geq \int_{z \in [-1/2, 1/2]^3} \left(\Psi, \sum_{\sigma \in \mathbb{Z}^3} \sum_{i=1}^N \left(\mathcal{K}_i^{(z+\sigma)} - \varepsilon \Delta_{i, \text{Neu}}^{(z+\sigma)} \right) \Psi \right) + T(S_z^\Psi) dz, -CL^3/\ell^5,$$

where $-\Delta_{\text{Neu}}^{(z)}$ is the Neumann Laplacian for the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$ and $S_z^\Psi : \mathbb{Z}^3 \rightarrow \mathbb{R}$ is the map

$$S_z^\Psi(\sigma) = \ell^{-1} \left(\sqrt{\widehat{n}_{z+\sigma}^+ + \widehat{n}_{z+\sigma}^-} + 1 \right) = \ell^{-1} \left(\sqrt{\widehat{n}_{z+\sigma}} + 1 \right), \quad (18)$$

where

$$\langle \widehat{n}_{z+\sigma}^+ + \widehat{n}_{z+\sigma}^- \rangle = \langle \Psi, (\widehat{n}_{z+\sigma}^+ + \widehat{n}_{z+\sigma}^-) \Psi \rangle.$$

In order to arrive at this lemma from Theorem B.1 we have used the inequality

$$T(S_z^\Psi) \leq \langle T(S_z^+) + T(S_z^-) \rangle,$$

where $S_z^\pm(\sigma) = \ell^{-1} \sqrt{\widehat{n}_{z+\sigma}^\pm} + 1/2$. To prove this inequality first note that by the 2-dimensional triangle inequality

$$\left(\sqrt{S_1^2 + \widetilde{S}_1^2} - \sqrt{S_2^2 + \widetilde{S}_2^2} \right)^2 \leq (S_1 - S_2)^2 + (\widetilde{S}_1 - \widetilde{S}_2)^2, \quad (19)$$

for all $S_1, S_2, \widetilde{S}_1, \widetilde{S}_2 \in \mathbb{R}$. The estimate (19) implies, in particular, that

$$n \mapsto T(\ell^{-1} \sqrt{n+1} - 1)$$

is a convex map on non-negative functions $n : \mathbb{Z}^3 \rightarrow \mathbb{R}$. Hence

$$T(S_z^\Psi) = T(\ell^{-1} \sqrt{\widehat{n}_{z+\sigma}} + 1 - 1) \leq \langle T(\ell^{-1} \sqrt{\widehat{n}_{z+\sigma}} + 1 - 1) \rangle \leq \langle T(S_z^+) + T(S_z^-) \rangle,$$

where the last inequality is again a consequence of (19).

As before we use the sliding technique to localize the electrostatic energy. If we combine Lemma 4.1 and Lemma 5.1 we obtain for $\Psi \in \bigotimes_S^N L^2(\mathbb{R}^3 \times \{1, -1\})$ normalized and with support in $([-L/2, L/2]^3 \times \{1, -1\})^N$ that

$$(\Psi, H_{N,L} \Psi) \geq \gamma \widetilde{\gamma} \int_{z \in [-1/2, 1/2]^3} \left(\Psi, \left(\sum_{\sigma \in \mathbb{Z}^3} \widetilde{H}_{z+\sigma} + \frac{\gamma_{\varepsilon,t}}{2} T(S_z) \right) \Psi \right) dz - CL^3/\ell^5 - \omega(t)N \left(\frac{1}{2\ell} - \frac{1}{2L} \right), \quad (20)$$

Ground State Energy of the Two-Component Charged Bose Gas

where

$$\gamma_{\varepsilon,t} = (1 + \varepsilon + Ct^3)^{-1}(\tilde{\gamma}\gamma)^{-1} \quad (21)$$

(recall that γ and $\tilde{\gamma}$ also depend on t) and for $z \in \mathbb{R}^3$ we have introduced the operator

$$\tilde{H}_z = \sum_{i=1}^N \left(\frac{\gamma_{\varepsilon,t}}{2} \mathcal{K}_i^{(z)} - \frac{\gamma_{\varepsilon,t}}{2} \varepsilon \Delta_{i,\text{Neu}}^{(z)} \right) + \sum_{1 \leq i < j \leq N} e_i e_j w_z(x_i, x_j), \quad (22)$$

with

$$w_z(x_i, x_j) = \chi_z(x_i) Y_{\frac{\omega(t)}{\ell}}(x_i - x_j) \chi_z(x_j). \quad (23)$$

We shall choose ε depending on N in Sect. 13 in such a way that $\varepsilon \rightarrow 0$ as $N \rightarrow \infty$. We shall in particular assume that $\varepsilon < 1$. Hence we may assume that $\gamma_{\varepsilon,t}$ is bounded above and below by constants.

The expressions $(\Psi, \sum_{\sigma \in \mathbb{Z}^3} \tilde{H}_{z+\sigma} \Psi) + \frac{1}{2} \gamma_{\varepsilon,t} T(S_z^\Psi)$ are equivalent for the different $z \in [-1/2, 1/2]^3$. It is therefore enough to consider $z = 0$. Moreover, the operator $\sum_{\sigma \in \mathbb{Z}^3} \tilde{H}_\sigma$ commutes with the number operators $\hat{v}_\sigma^\pm$ (the operators giving the number of positively and negatively charged particles in the cube $\{\ell\sigma\} + [-\ell/2, \ell/2]^3$). Note that if Ψ has support in the set $([-L/2, L/2]^3 \times \{1, -1\})^N$ then for z outside the set $\Omega = [-\frac{L}{2\ell} - \frac{1}{2}, \frac{L}{2\ell} + \frac{1}{2}]^3$ we have that Ψ is in the kernel of the operators $\hat{v}_z^\pm$.

The estimate (20) then implies the following result.

Theorem 5.2 (Localization into small cubes). *With the definitions from (10), (13–18), and (21–23) we have that*

$$\begin{aligned} E_L(N) &\geq \gamma \tilde{\gamma} \inf_{\Psi, \|\Psi\|=1} \left(\left(\Psi, \sum_{\sigma \in \mathbb{Z}^3} \tilde{H}_\sigma \Psi \right) + \frac{\gamma_{\varepsilon,t}}{2} T(S_0^\Psi) \right) \\ &\quad - CL^3/\ell^5 - \omega(t)N \left(\frac{1}{2\ell} - \frac{1}{2L} \right), \end{aligned}$$

where the infimum is over normalized functions

$$\begin{aligned} \Psi \in \mathcal{H}_0 &= \left\{ \Psi \in \bigotimes_s^N L^2(\mathbb{R}^3 \times [-1, 1]) \mid \hat{v}_\sigma^\pm \Psi \right. \\ &\quad \left. = 0 \text{ for } \sigma \notin \mathbb{Z}^3 \cap \left[-\frac{L}{2\ell} - \frac{1}{2}, \frac{L}{2\ell} + \frac{1}{2} \right]^3 \right\}, \end{aligned}$$

and $0 < t < 1/2$ is the parameter from the beginning of Sect. 4.

In the following Sects. 6–11 we shall study the operators $\tilde{H}_\sigma$, $\sigma \in \mathbb{Z}^3$. Since $\tilde{H}_\sigma$ commutes with $\hat{v}_\sigma^\pm$ we may simply restrict to the eigenspaces of $\hat{v}_\sigma^\pm$. We shall therefore not think of $\hat{v}_\sigma^\pm$ and $\hat{v}_\sigma$ as operators, but as classical parameters $v_\sigma^\pm$ and v_σ . Which values of these parameters that will give the optimal energy is of course not known a-priori, but they must satisfy $\sum_\sigma v_\sigma = N$. The operator $\tilde{H}_\sigma$ is in (22) written as an N -particle operator, but it depends only on v_σ particles. We shall therefore simply think of it as a v_σ -particle operator. The operators $\tilde{H}_\sigma$ for different σ are all unitarily equivalent. We shall in Sects. 6–11 simply omit the subscript σ .

6. Long and Short Distance Cutoffs in the Potential

Our aim in this section is to replace w in (23) (omitting the index z) by a function that has long and short distance cutoffs.

We shall replace the function w by

$$w_{r,R}(x, y) = \chi(x) V_{r,R}(x - y) \chi(y), \quad (24)$$

where

$$V_{r,R}(x) = Y_{R-1}(x) - Y_{r-1}(x) = \frac{e^{-|x|/R} - e^{-|x|/r}}{|x|}. \quad (25)$$

Here $0 < r \leq R \leq \omega(t)^{-1}\ell$. Note that for $x \ll r$ then $V_{r,R}(x) \approx r^{-1} - R^{-1}$ and for $|x| \gg R$ then $V_{r,R}(x) \approx |x|^{-1}e^{-|x|/R}$.

We first introduce the cutoff R alone, i.e., we bound the effect of replacing w by $w_R(x, y) = \chi(x) V_R(x - y) \chi(y)$, where $V_R(x) = |x|^{-1}e^{-|x|/R} = Y_{R-1}(x)$. Thus, since $R \leq \omega(t)^{-1}\ell$, the Fourier transforms satisfy

$$\widehat{Y}_{\omega/\ell}(k) - \widehat{V}_R(k) = 4\pi \left(\frac{1}{k^2 + (\omega(t)/\ell)^2} - \frac{1}{k^2 + R^{-2}} \right) \geq 0.$$

(We use the convention that $\hat{f}(k) = \int f(x) e^{-ikx} dx$.) Hence

$$w(x, y) - w_R(x, y) = \chi(x) (Y_{\omega/\ell} - V_R)(x - y) \chi(y)$$

defines a positive semi-definite kernel. Note, moreover, that $(Y_{\omega/\ell} - V_R)(0) = R^{-1} - \omega/\ell \leq R^{-1}$. Thus,

$$\begin{aligned} & \sum_{1 \leq i < j \leq N} e_i e_j (w(x_i, x_j) - w_R(x_i, x_j)) \\ &= \frac{1}{2} \sum_{1 \leq i, j \leq N} e_i e_j (w - w_R)(x_i, x_j) - \sum_{i=1}^N \frac{e_i^2}{2} (w - w_R)(x_i, x_i) \\ &\geq -\frac{1}{2} v (Y_{\omega/\ell} - V_R)(0) \geq -\frac{1}{2} v R^{-1}. \end{aligned} \quad (26)$$

Here we have used that $\sum_{i=1}^N \chi(x_i)^2 \leq \sum_{i=1}^N X(x_i) = v$.

We shall now bound the effect of replacing w_R by $w_{r,R}$. I.e., we are replacing $V_R(x) = |x|^{-1}e^{-|x|/R}$ by $|x|^{-1}(e^{-|x|/R} - e^{-|x|/r})$. The correction is $Y_{r-1}(x) = |x|^{-1}e^{-|x|/r}$.

Lemma 6.1. *We have for all $\delta > 0$ the operator inequality*

$$\begin{aligned} & \sum_{i=1}^N -\frac{\delta}{2} \Delta_{i, \text{Neu}}^{(z)} + \sum_{1 \leq i < j \leq N} e_i e_j \chi(x_i) Y_{r-1}(x_i - x_j) \chi(x_j) \\ &\geq -C v^+ v^- (r^2 \ell^{-3} + \delta^{-3/2} r^{1/2}). \end{aligned}$$

Ground State Energy of the Two-Component Charged Bose Gas

Proof. We set $z = 0$ and omit the index z in this proof. For $D > 0$ we define

$$W_D(x) = \begin{cases} Y_{r-1}(D), & |x| < D \\ Y_{r-1}(x), & |x| \geq D \end{cases}.$$

Then the measure

$$\begin{aligned} \mu_D &= (4\pi)^{-1}(-\Delta + r^{-2})W_D \\ &= (4\pi)^{-1}r^{-2}Y_{r-1}(D)\mathbf{1}_{|x|<D} - (4\pi)^{-1}Y'_{r-1}(D)\delta(|x| - D) \end{aligned}$$

is non-negative and $W_D = \mu_D * Y_{r-1}$. Here Y'_{r-1} refers to the radial derivative of Y_{r-1} and $\mathbf{1}_{|x|<D}$ is the characteristic function of the set $\{|x| < D\}$.

Define the (signed) measure

$$d\mu(x) = \sum_{i=1}^N e_i \chi(x_i) d\mu_{D_i}(x - x_i),$$

where

$$D_i = \frac{1}{2} \min \left\{ |x_i - x_j| \mid j = 1, \dots, N, e_i e_j = -1, x_i \in [-\ell/2, \ell/2]^3 \right\},$$

is half the distance from the particle at x_i to the nearest particle of opposite charge in the cube $[-\ell/2, \ell/2]^3$, i.e., the cube in which χ is supported. Thus for $x_i, x_j \in [-\ell/2, \ell/2]^3$ we have

$$\begin{aligned} \iint Y_{r-1}(x - y) d\mu_{D_i}(x - x_i) d\mu_{D_j}(y - x_j) &= \int W_{D_j}(x - x_j) d\mu_{D_i}(x - x_i) \\ &\leq \int Y_{r-1}(x - x_j) d\mu_{D_i}(x - x_i) \\ &= W_{D_i}(x_i - x_j) \leq Y_{r-1}(x_i - x_j) \end{aligned}$$

and both inequalities become equalities if $e_i e_j = -1$. Hence

$$\begin{aligned} &\sum_{1 \leq i < j \leq N} e_i e_j \chi(x_i) Y_{r-1}(x_i - x_j) \chi(x_j) \\ &\geq \frac{1}{2} \iint Y_{r-1}(x - y) d\mu(x) d\mu(y) \\ &\quad - \frac{1}{2} \sum_{i=1}^N \chi(x_i)^2 \iint Y_{r-1}(x - y) d\mu_{D_i}(x) d\mu_{D_i}(y) \\ &\geq -\frac{1}{2} \sum_{i=1}^N \chi(x_i)^2 \iint Y_{r-1}(x - y) d\mu_{D_i}(x) d\mu_{D_i}(y) \end{aligned}$$

since Y_{r-1} is positive type. This inequality is very similar to Onsager's electrostatic inequality [O].

We calculate

$$\begin{aligned} \iint Y_{r^{-1}}(x-y) d\mu_D(x) d\mu_D(y) &= \int W_D(x) d\mu_D(x) \\ &= \frac{1}{3} r^{-2} D^3 Y_{r^{-1}}(D)^2 - D^2 Y'_{r^{-1}}(D) Y_{r^{-1}}(D) \\ &= \frac{1}{3} \left(D r^{-2} + 3 D^{-1} + 3 r^{-1} \right) e^{-2D/r}. \end{aligned}$$

We have therefore proved the operator estimate

$$\sum_{i=1}^N -\frac{\delta}{2} \Delta_{i,\text{Neu}} + \sum_{1 \leq i < j \leq N} e_i e_j \chi(x_i) Y_{r^{-1}}(x_i - x_j) \chi(x_j) \geq \sum_{i=1}^N -\frac{\delta}{2} \Delta_{i,\text{Neu}} - v_i,$$

where for $i = 1, 2, \dots, N$

$$v_i(x_i) = -\frac{1}{6} \chi(x_i)^2 \left(D_i r^{-2} + 3 D_i^{-1} + 3 r^{-1} \right) e^{-2D_i/r}.$$

Note that here D_i depends on x_i and the positions of all the particles with opposite charge of e_i .

From the Sobolev estimate for the Neumann Laplacian in a cube of size ℓ we have the general lower bound

$$-\Delta_{\text{Neu}} - V \geq -C \int V^{5/2} - C \ell^{-3} \int V.$$

This gives for $e_i = \pm 1$ that

$$\begin{aligned} -\frac{\delta}{2} \Delta_{i,\text{Neu}} - v_i &\geq -C v^\mp \int_{\mathbb{R}^3} \delta^{-3/2} \left(\frac{|x|}{r^2} + \frac{3}{|x|} + \frac{3}{r} \right)^{5/2} e^{-5|x|/r} \\ &\quad + \ell^{-3} \left(\frac{|x|}{r^2} + \frac{3}{|x|} + \frac{3}{r} \right) e^{-2|x|/r} dx \\ &\geq -C v^\mp (\delta^{-3/2} r^{1/2} + \ell^{-3} r^2). \end{aligned}$$

When summed over i this gives the result of the lemma. $\square$

If we combine the bound (26) and Lemma 6.1 we have the following result.

Lemma 6.2 (Long and short distance potential cutoffs). *For $0 < \delta < \varepsilon < 1$ consider the Hamiltonian*

$$\tilde{H}_{r,R}^\delta = \sum_{i=1}^N \left(\frac{\gamma_{\varepsilon,t}}{2} \mathcal{K}_i^{(z)} - \frac{\gamma_{\varepsilon,t}}{2} (\varepsilon - \delta) \Delta_{i,\text{Neu}}^{(z)} \right) + \sum_{1 \leq i < j \leq N} e_i e_j w_{r,R}(x_i, x_j), \quad (27)$$

where $w_{r,R}$ is given in (24) and (25). If $0 < r \leq R < \omega(t)^{-1} \ell$ then the Hamiltonian $\tilde{H}$ defined in (22) obeys the lower bound

$$\tilde{H} \geq \tilde{H}_{r,R}^\delta - \frac{1}{2} v R^{-1} - C v^+ v^- (\delta^{-3/2} r^{1/2} + \ell^{-3} r^2). \quad (28)$$

If $0 < r \leq R = \omega(t)^{-1} \ell$ we get

$$\tilde{H} \geq \tilde{H}_{r,R}^\delta - C v^+ v^- (\delta^{-3/2} r^{1/2} + \ell^{-3} r^2). \quad (29)$$

A similar argument gives the following result.

Lemma 6.3. *With the same notation as above we have for $0 < r \leq r' \leq R' \leq R \leq \omega(t)^{-1}\ell$ that*

$$\tilde{H}_{r,R}^\delta \geq \tilde{H}_{r',R'}^{2\delta} - \frac{1}{2}\nu R'^{-1} - C\nu^+ \nu^- (\delta^{-3/2} r'^{1/2} + \ell^{-3} r'^2).$$

Proof. Simply note that $V_{r',R'}(x) - V_{r,R}(x) = Y_{R'-1}(x) - Y_{R-1}(x) + Y_{r-1}(x) - Y_{r'-1}(x)$ and now use the same arguments as before. $\square$

We shall not fix the cutoffs r and R nor the parameter δ , but rather choose them differently at different stages in the later arguments. Moreover, we shall choose r and R to depend on ν . We give an example of this in the following corollary.

Corollary 6.4 (Simple bound on the number of excited particles). *For any state such that the expectation $\langle \tilde{H} \rangle \leq 0$, the expectation of the number of excited particles satisfies $\langle \nu - \hat{n} \rangle \leq \nu \min\{1, C\varepsilon^{-1}\ell\nu^{1/3}\}$. (Recall that $\tilde{H}$ depends on ε .)*

Proof. We simply choose $r = R$ and $\delta = \varepsilon/2$ in Lemma 6.2. Then $\tilde{H}_{r,R}^\delta \geq C(\varepsilon/2)(\nu - \hat{n})\ell^{-2}$ since $(u, -\Delta_{\text{Neu}}^{(z)} u) \geq \pi^2 \ell^{-2} \|u\|^2$ for functions u in the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$ orthogonal to constants. We therefore have that

$$\tilde{H} \geq C(\varepsilon/2)(\nu - \hat{n})\ell^{-2} - \frac{1}{2}\nu r^{-1} - C\nu^+ \nu^- (\varepsilon^{-3/2} r^{1/2} + \ell^{-3} r^2).$$

Strictly speaking Lemma 6.2 requires $r \leq R \leq \omega(t)^{-1}\ell$. If however $R = r > \omega(t)^{-1}\ell$ we would get an even better estimate than the one stated if we set $r = R = \omega(t)^{-1}\ell$ and $\delta = \varepsilon/2$. Since in this case we may ignore the error $-\frac{1}{2}\nu r^{-1}$.

We now estimate $\nu^+ \nu^- \leq \nu^2/4$ and make the explicit choice $r = \min\{\varepsilon\nu^{-2/3}, \ell\nu^{-1/3}\}$. When $\varepsilon\nu^{-2/3} \leq \ell\nu^{-1/3}$, i.e., $\nu^{1/3} \geq \varepsilon\ell^{-1}$ we obtain

$$\begin{aligned} \tilde{H} &\geq C(\varepsilon/2)(\nu - \hat{n})\ell^{-2} - C\varepsilon^{-1}\nu^{5/3} - C\varepsilon^2\ell^{-3}\nu^{4/3} \\ &\geq C(\varepsilon/2)(\nu - \hat{n})\ell^{-2} - C\varepsilon^{-1}\nu^{5/3} \end{aligned}$$

and when $\varepsilon\nu^{-2/3} \geq \ell\nu^{-1/3}$, i.e., $\nu^{1/3} \leq \varepsilon\ell^{-1}$ we obtain

$$\begin{aligned} \tilde{H} &\geq C(\varepsilon/2)(\nu - \hat{n})\ell^{-2} - C\varepsilon^{-3/2}\ell^{1/2}\nu^{11/6} - C\ell^{-1}\nu^{4/3} \\ &\geq C(\varepsilon/2)(\nu - \hat{n})\ell^{-2} - C\ell^{-1}\nu^{4/3}. \end{aligned}$$

These two bounds give that

$$\langle \nu - \hat{n} \rangle \leq C\nu \max\{(\varepsilon^{-1}\ell\nu^{1/3})^2, \varepsilon^{-1}\ell\nu^{1/3}\}.$$

We of course clearly have that $\nu - \hat{n} \leq \nu$. We therefore get the result claimed. $\square$

7. Bound on the Unimportant Part of the Hamiltonian

In this section we shall bound the Hamiltonian $\tilde{H}_{r,R}^\delta$ given in (27). We emphasize that we do not necessarily have neutrality in the cube, i.e., ν^+ and ν^- may be different. We are simply looking for a lower bound to $\tilde{H}_{r,R}^\delta$, that holds for all $\nu^\pm$. The goal is to find a lower bound that will allow us to conclude that the optimal value of $\nu^+ - \nu^-$, i.e., the value for which the energy of the Hamiltonian is smallest, is indeed close to zero. We

shall also conclude that the number of excited particles $v - \hat{n}$ is small. These conclusions will be made in Sects. 9 and 10.

In this section we shall fix $z = 0$, i.e., we are working in the cube $[-\ell/2, \ell/2]^3$. We shall simply omit the index z .

We shall express the Hamiltonian in second quantized language. This is purely for convenience. We stress that we are not in anyway changing the model by doing this and the treatment is entirely rigorous and could have been done without the use of second quantization.

Let $u_p, \ell p/\pi \in (\mathbb{N} \cup \{0\})^3$ be an orthonormal basis of eigenfunction of the Neumann Laplacian $-\Delta_{\text{Neu}}$ in the cube $[-\ell/2, \ell/2]^3$. More precisely, $-\Delta_{\text{Neu}} u_p = |p|^2 u_p$. I.e.,

$$u_p(x_1, x_2, x_3) = c_p \ell^{-3/2} \prod_{j=1}^3 \cos(p_j(x_j + \ell/2)),$$

where the normalization satisfies $c_0 = 1$ and in general $1 \leq c_p \leq \sqrt{8}$. The function $u_0 = \ell^{-3/2}$ is the constant eigenfunction with eigenvalue 0. We note that for $p \neq 0$ we have

$$(u_p, -\Delta_{\text{Neu}} u_p) \geq \pi^2 \ell^{-2}. \quad (30)$$

We now express the Hamiltonian $\tilde{H}_{r,R}^\delta$ from (27) (omitting z) in terms of the creation and annihilation operators $a_{p\pm}^* = a_\pm(u_p)^*$ and $a_{p\pm} = a_\pm(u_p)$.

Define

$$\hat{w}_{pq,\mu\nu} = \iint w_{r,R}(x, y) u_p(x) u_q(y) u_\mu(x) u_\nu(y) dx dy.$$

We may then express the two-body potential in $\tilde{H}_{r,R}^\delta$ as

$$\begin{aligned} & \sum_{1 \leq i < j \leq n} e_i e_j w_{r,R}(x_i, x_j) \\ &= \frac{1}{2} \sum_{pq,\mu\nu} \hat{w}_{pq,\mu\nu} \left(a_{p+}^* a_{q+}^* a_{v+} a_{\mu+} + a_{p-}^* a_{q-}^* a_{v-} a_{\mu-} - 2a_{p+}^* a_{q-}^* a_{v-} a_{\mu+} \right). \end{aligned}$$

Motivated by Foldy's and Dyson's use of the Bogolubov approximation it is our goal to reduce the Hamiltonian $\tilde{H}_{r,R}^\delta$ so that it has only what we call quadratic terms, i.e., terms which contain precisely two $a_{p\pm}^\#$ with $p \neq 0$. More precisely, we want to be able to ignore all terms in the two-body-potential containing the coefficients

- $\hat{w}_{00,00}$.
- $\hat{w}_{p0,q0} = \hat{w}_{0p,0q}$, where $p, q \neq 0$. These terms are in fact quadratic, but do not appear in the Foldy Hamiltonian. We shall prove that they can also be ignored.
- $\hat{w}_{p0,00} = \hat{w}_{0p,00} = \hat{w}_{00,p0} = \hat{w}_{00,0p}$, where $p \neq 0$.
- $\hat{w}_{pq,\mu 0} = \hat{w}_{\mu 0,pq} = \hat{w}_{qp,0\mu} = \hat{w}_{0\mu,qp}$, where $p, q, \mu \neq 0$.
- $\hat{w}_{pq,\mu\nu}$, where $p, q, \mu, \nu \neq 0$.

We shall consider these cases one at a time.

Ground State Energy of the Two-Component Charged Bose Gas

Lemma 7.1 (Control of terms with $\widehat{w}_{00,00}$). *The sum of the terms in $\widetilde{H}_{r,R}^\delta$ containing $\widehat{w}_{00,00}$ is equal to*

$$\frac{1}{2}\widehat{w}_{00,00}\left[(\widehat{n}^+ - \widehat{n}^-)^2 - \widehat{n}^+ - \widehat{n}^-\right].$$

Proof. The terms containing $\widehat{w}_{00,00}$ are

$$\frac{1}{2}\widehat{w}_{00,00}\left(a_{0+}^*a_{0+}^*a_{0+}a_{0+} + a_{0-}^*a_{0-}^*a_{0-}a_{0-} - 2a_{0+}^*a_{0-}^*a_{0-}a_{0+}\right),$$

which gives the above result when using the commutation relation $[a_p, a_q^*] = \delta_{p,q}$. $\square$

Lemma 7.2 (Control of terms with $\widehat{w}_{p0,q0}$). *The sum of the terms in $\widetilde{H}_{r,R}^\delta$ containing $\widehat{w}_{p0,q0}$ or $\widehat{w}_{0p,0q}$ with $p, q \neq 0$ is bounded below by*

$$-4\pi R^2 \ell^{-3} |\widehat{n}^+ - \widehat{n}^-| (v - \widehat{n}).$$

Proof. The terms containing $\widehat{w}_{p0,q0} = \widehat{w}_{0p,0q}$ are

$$\begin{aligned} & \sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} \left(a_{p+}^* a_{0+}^* a_{0+} a_{q+} + a_{p-}^* a_{0-}^* a_{0-} a_{q-} - a_{p+}^* a_{0-}^* a_{0-} a_{q+} - a_{p-}^* a_{0+}^* a_{0+} a_{q-} \right) \\ &= \sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} (\widehat{n}^+ - \widehat{n}^-) (a_{p+}^* a_{q+} - a_{p-}^* a_{q-}). \end{aligned}$$

Note that $\widehat{n}^\pm$ commutes with $\sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} a_{p\pm}^* a_{q\pm}$.

We have that

$$\widehat{w}_{p0,q0} = \ell^{-3} \int \int w_{r,R}(x, y) dy u_p(x) u_q(x) dx.$$

Hence

$$\begin{aligned} & \sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} a_{p\pm}^* a_{q\pm} \\ &= \ell^{-3} \int \int w_{r,R}(x, y) dy \left(\sum_{p \neq 0} u_p(x) a_{p\pm}^* \right) \left(\sum_{p \neq 0} u_p(x) a_{p\pm}^* \right)^* dx. \\ &\leq \ell^{-3} \sup_{x'} \int w_{r,R}(x', y) dy \int \left(\sum_{p \neq 0} u_p(x) a_{p\pm}^* \right) \left(\sum_{p \neq 0} u_p(x) a_{p\pm}^* \right)^* dx. \\ &= \ell^{-3} \sup_{x'} \int w_{r,R}(x', y) dy \sum_{p \neq 0} a_{p\pm}^* a_{p\pm}. \end{aligned}$$

Since $\sum_{p \neq 0} a_{p\pm}^* a_{p\pm} = v^\pm - \widehat{n}^\pm$ and

$$\sup_x \int w_{r,R}(x, y) dy \leq \int V_{r,R}(y) dy \leq 4\pi R^2$$

we obtain the operator inequality

$$0 \leq \sum_{\substack{p \neq 0 \\ q \neq 0}} \widehat{w}_{p0,q0} a_{p\pm}^* a_{q\pm} \leq 4\pi \ell^{-3} R^2 (v^\pm - \widehat{n}^\pm),$$

and the lemma follows. $\square$

Before treating the last three types of terms we shall need the following result on the structure of the coefficients $\widehat{w}_{pq,\mu\nu}$.

Lemma 7.3. *For all $p', q' \in (\pi/\ell) (\mathbb{N} \cup \{0\})^3$ and $\alpha \in \mathbb{N}$ there exists $J_{p'q'}^\alpha \in \mathbb{R}$ with $J_{p'q'}^\alpha = J_{q'p'}^\alpha$ such that for all $p, q, \mu, \nu \in (\pi/\ell) (\mathbb{N} \cup \{0\})^3$ we have*

$$\widehat{w}_{pq,\mu\nu} = \sum_{\alpha} J_{p\mu}^\alpha J_{q\nu}^\alpha. \quad (31)$$

Moreover we have the operator inequalities

$$0 \leq \sum_{p,p' \neq 0} \widehat{w}_{pp',00} a_{p\pm}^* a_{p'\pm} = \sum_{p,p' \neq 0} \widehat{w}_{p0,0p'} a_{p\pm}^* a_{p'\pm} \leq 4\pi \ell^{-3} R^2 (v^\pm - \widehat{n}^\pm) \quad (32)$$

and

$$0 \leq \sum_{p,p',m \neq 0} \widehat{w}_{pm,mp'} a_{p\pm}^* a_{p'\pm} \leq r^{-1} (v^\pm - \widehat{n}^\pm).$$

Proof. The operator $\mathcal{A}$ with integral kernel $w_{r,R}(x, y)$ is a non-negative Hilbert-Schmidt operator on $L^2(\mathbb{R}^3)$ with norm less than $\sup_k \widehat{V}_{r,R}(k) \leq 4\pi R^2$. Denote the eigenvalues of $\mathcal{A}$ by λ_α , $\alpha = 1, 2, \dots$ and corresponding orthonormal eigenfunctions by φ_α . We may assume that these functions are real. The eigenvalues satisfy $0 \leq \lambda_\alpha \leq 4\pi R^2$. We then have

$$\widehat{w}_{pq,\mu\nu} = \sum_{\alpha} \lambda_{\alpha} \int u_p(x) u_{\mu}(x) \varphi_{\alpha}(x) dx \int u_q(y) u_{\nu}(y) \varphi_{\alpha}(y) dy.$$

The identity (31) thus follows with $J_{p\mu}^\alpha = \lambda_{\alpha}^{1/2} \int u_p(x) u_{\mu}(x) \varphi_{\alpha}(x) dx$.

If as before $\mathcal{P}$ denotes the projection orthogonal to constants we may also consider the operator $\mathcal{PAP}$. Denote its eigenvalues and eigenfunctions by λ'_{α} and φ'_{α} . Then again $0 \leq \lambda'_{\alpha} \leq 4\pi R^2$. Hence we may write

$$\widehat{w}_{p0,0p'} = \ell^{-3} \sum_{\alpha} \lambda'_{\alpha} \int u_p(x) \varphi'_{\alpha}(x) dx \int u_{p'}(y) \varphi'_{\alpha}(y) dy.$$

Thus since all φ'_{α} are orthogonal to constants we have

$$\begin{aligned} & \sum_{p,p' \neq 0} \widehat{w}_{p0,0p'} a_{p\pm}^* a_{p'\pm} \\ &= \ell^{-3} \sum_{\alpha} \lambda'_{\alpha} \left(\sum_{p \neq 0} \int u_p(x) \varphi'_{\alpha}(x) dx a_{p\pm}^* \right) \left(\sum_{p' \neq 0} \int u_{p'}(x) \varphi'_{\alpha}(x) dx a_{p\pm}^* \right)^* \\ &= \ell^{-3} \sum_{\alpha} \lambda'_{\alpha} a_{\pm}^* (\varphi'_{\alpha}) a_{\pm} (\varphi'_{\alpha}). \end{aligned}$$

The inequalities (32) follow immediately from this.

Ground State Energy of the Two-Component Charged Bose Gas

The fact that $\sum_{p,p',m \neq 0} \widehat{w}_{pm,mp'} a_{p\pm}^* a_{p'\pm} \geq 0$ follows from the representation (31). Moreover, since the kernel $w_{r,R}(x, y)$ is a continuous function we have that $w_{r,R}(x, x) = \sum_{\alpha} \lambda_{\alpha} \varphi_{\alpha}(x)^2$ for almost all x and hence

$$\sum_{m \neq 0} \widehat{w}_{pm,mp'} = \int u_p(x) u_{p'}(x) w_{r,R}(x, x) dx - \widehat{w}_{p0,0p'}.$$

We therefore have

$$\begin{aligned} \sum_{p,p',m \neq 0} \widehat{w}_{pm,mp'} a_{p\pm}^* a_{p'\pm} &\leq \sum_{p,p' \neq 0} \int u_p(x) u_{p'}(x) w_{r,R}(x, x) dx a_{p\pm}^* a_{p'\pm} \\ &= \int w_{r,R}(x, x) \left(\sum_{p \neq 0} u_p(x) a_{p\pm}^* \right) \left(\sum_{p \neq 0} u_p(x) a_{p\pm} \right)^* dx \\ &\leq \sup_{x'} w_{r,R}(x', x') \int \left(\sum_{p \neq 0} u_p(x) a_{p\pm}^* \right) \left(\sum_{p \neq 0} u_p(x) a_{p\pm} \right)^* dx \\ &= \sup_{x'} w_{r,R}(x', x') (v^{\pm} - \widehat{n}^{\pm}) \end{aligned}$$

and the lemma follows since $\sup_{x'} w_{r,R}(x', x') \leq r^{-1}$. $\square$

Lemma 7.4 (Control of terms with $\widehat{w}_{pq,\mu\nu}$). *The sum of the terms in $\widetilde{H}_{r,R}^{\delta}$ containing $\widehat{w}_{pq,\mu\nu}$, with $p, q, \mu, \nu \neq 0$ is bounded below by*

$$\frac{1}{2} \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2 - \frac{1}{2} r^{-1} (v - \widehat{n}).$$

Proof. The relevant terms are

$$\begin{aligned} &\frac{1}{2} \sum_{pq,\mu\nu \neq 0} \widehat{w}_{pq,\mu\nu} \left(a_{p+}^* a_{q+}^* a_{\nu+} a_{\mu+} + a_{p-}^* a_{q-}^* a_{\nu-} a_{\mu-} - 2a_{p+}^* a_{q-}^* a_{\nu-} a_{\mu+} \right) \\ &= \frac{1}{2} \sum_{pq,\mu\nu \neq 0} \widehat{w}_{pq,\mu\nu} \left(a_{p+}^* a_{\mu+} a_{q+}^* a_{\nu+} + a_{p-}^* a_{\mu-} a_{q-}^* a_{\nu-} - 2a_{p+}^* a_{q-}^* a_{\nu-} a_{\mu+} \right) \\ &\quad - \frac{1}{2} \sum_{pp',m \neq 0} \widehat{w}_{pm,mp'} a_{p+}^* a_{p'+} - \frac{1}{2} \sum_{pp',m \neq 0} \widehat{w}_{pm,mp'} a_{p-}^* a_{p'-} \\ &= \frac{1}{2} \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2 - \frac{1}{2} r^{-1} (v - \widehat{n}), \end{aligned}$$

where we have used the last estimate in Lemma 7.3 and that $v^{+} - \widehat{n}^{+} + v^{-} - \widehat{n}^{-} = v - \widehat{n}$. $\square$

Lemma 7.5 (Control of terms with $\widehat{w}_{p0,00}$). *The sum of the terms in $\widetilde{H}_{r,R}^{\delta}$ containing $\widehat{w}_{p0,00}$, $\widehat{w}_{0p,00}$, $\widehat{w}_{00,p0}$, or $\widehat{w}_{00,0p}$, with $p \neq 0$ is, for all $\varepsilon' > 0$, bounded below by*

$$-\varepsilon'^{-1} 4\pi \ell^{-3} R^2 (\widehat{n} + 1) (v - \widehat{n}) - 2\varepsilon' \widehat{w}_{00,00} \left((\widehat{n}^{+} - \widehat{n}^{-})^2 + 1 \right), \quad (33)$$

or by

$$(v^+ - v^-) \sum_{p \neq 0} \widehat{w}_{p0,00} \left(a_{p^+}^* a_{0^+} - a_{p^-}^* a_{0^-} + a_{0^+}^* a_{p^+} - a_{0^-}^* a_{p^-} \right) - \varepsilon'^{-1} 4\pi \ell^{-3} R^2 (\widehat{n}^+ + 1)(v - \widehat{n}) - 2\varepsilon' \widehat{w}_{00,00} \left((v - \widehat{n})^2 + 1 \right). \quad (34)$$

Proof. The terms containing $\widehat{w}_{p0,00}$, $\widehat{w}_{0p,00}$, $\widehat{w}_{00,p0}$, or $\widehat{w}_{00,0p}$ are

$$\begin{aligned} & \sum_{p \neq 0} \frac{1}{2} \widehat{w}_{p0,00} \left(2a_{p^+}^* a_{0^+} + a_{0^+} a_{0^+} + 2a_{0^+}^* a_{0^+} a_{p^+} + 2a_{p^-}^* a_{0^-} a_{0^-} \right. \\ & \quad \left. + 2a_{0^-}^* a_{0^-} a_{p^-} - 2a_{0^-}^* a_{0^-} a_{0^+} a_{p^+} \right. \\ & \quad \left. - 2a_{0^-}^* a_{0^-} a_{p^+} a_{0^+} - 2a_{0^+}^* a_{0^+} a_{p^-} a_{0^-} - 2a_{0^+}^* a_{0^+} a_{p^-} a_{0^-} \right) \\ &= \sum_{\alpha} \sum_{p \neq 0} J_{p0}^{\alpha} J_{00}^{\alpha} \left((\widehat{n}^+ - \widehat{n}^-) (a_{p^+}^* a_{0^+} - a_{p^-}^* a_{0^-}) + (a_{0^+}^* a_{p^+} - a_{0^-}^* a_{p^-}) (\widehat{n}^+ - \widehat{n}^-) \right) \\ &= \sum_{\alpha} \sum_{p \neq 0} J_{p0}^{\alpha} J_{00}^{\alpha} \left(a_{p^+}^* a_{0^+} (\widehat{n}^+ + 1 - \widehat{n}^-) + (\widehat{n}^+ + 1 - \widehat{n}^-) a_{0^+}^* a_{p^+} \right) \\ & \quad - \sum_{\alpha} \sum_{p \neq 0} J_{p0}^{\alpha} J_{00}^{\alpha} \left(a_{p^-}^* a_{0^-} (\widehat{n}^+ - \widehat{n}^- - 1) + a_{0^-}^* a_{p^-} (\widehat{n}^+ - \widehat{n}^- - 1) \right) \\ &\geq -\varepsilon' \sum_{\alpha} \left(J_{00}^{\alpha} \right)^2 \left((\widehat{n}^+ - \widehat{n}^- + 1)^2 + (\widehat{n}^+ - \widehat{n}^- - 1)^2 \right) - \varepsilon'^{-1} (\widehat{n}^+ + 1) \\ & \quad \times \sum_{\alpha} \sum_{p, p' \neq 0} J_{p0}^{\alpha} J_{p'0}^{\alpha} a_{p^+}^* a_{p'^+} - \varepsilon'^{-1} (\widehat{n}^- + 1) \sum_{\alpha} \sum_{p, p' \neq 0} J_{p0}^{\alpha} J_{p'0}^{\alpha} a_{p^-}^* a_{p'^-}. \end{aligned}$$

Here we have used the representation (31) and in the last step a Cauchy-Schwarz inequality. We arrive at the first bound in the lemma since from (32) we have that

$$\sum_{\alpha} \sum_{p, p' \neq 0} J_{p0}^{\alpha} J_{p'0}^{\alpha} a_{p^{\pm}}^* a_{p'^{\pm}} = \sum_{p, p' \neq 0} \widehat{w}_{p0,0p'} a_{p^{\pm}}^* a_{p'^{\pm}} \leq 4\pi \ell^{-3} R^2 (v^{\pm} - \widehat{n}^{\pm}).$$

The second bound (34) follows in the same way if we notice that the terms containing $\widehat{w}_{p0,00}$, $\widehat{w}_{0p,00}$, $\widehat{w}_{00,p0}$, or $\widehat{w}_{00,0p}$ may be written as

$$\begin{aligned} & (v^+ - v^-) \sum_{p \neq 0} \widehat{w}_{p0,00} (a_{p^+}^* a_{0^+} - a_{p^-}^* a_{0^-} + a_{0^+}^* a_{p^+} - a_{0^-}^* a_{p^-}) \\ & \quad + \sum_{\alpha} \sum_{p \neq 0} J_{p0}^{\alpha} J_{00}^{\alpha} \left((\widehat{n}^+ - v^+ - \widehat{n}^- + v^-) (a_{p^+}^* a_{0^+} - a_{p^-}^* a_{0^-}) \right. \\ & \quad \left. + (a_{0^+}^* a_{p^+} - a_{0^-}^* a_{p^-}) (\widehat{n}^+ - v^+ - \widehat{n}^- + v^-) \right). \end{aligned}$$

□

Lemma 7.6 (Control of terms with $\widehat{w}_{pq,m0}$). *The sum of the terms in $\widetilde{H}_{r,R}^{\delta}$ containing $\widehat{w}_{pq,m0}$, $\widehat{w}_{qp,0m}$, $\widehat{w}_{m0,pq}$, or $\widehat{w}_{0m,qp}$, with $p, q, m \neq 0$ is bounded below by*

$$-\varepsilon'^{-1} 4\pi \ell^{-3} R^2 (\widehat{n} + 1)(v - \widehat{n}) - 2\varepsilon' \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p^+}^* a_{\mu^+} - a_{p^-}^* a_{\mu^-}) \right)^2,$$

for all $\varepsilon' > 0$.

Ground State Energy of the Two-Component Charged Bose Gas

Proof. The terms containing $\widehat{w}_{pq,m0}$, $\widehat{w}_{qp,0m}$, $\widehat{w}_{m0,pq}$, or $\widehat{w}_{0m,qp}$, with $p, q, m \neq 0$ are

$$\begin{aligned}
 & \sum_{pqm \neq 0} \widehat{w}_{pq,m0} \left(a_{p+}^* a_{q+}^* a_{m+} a_{0+} + a_{0+}^* a_{m+}^* a_{q+} a_{p+} + a_{p-}^* a_{q-}^* a_{m-} a_{0-} \right. \\
 & \quad \left. + a_{0-}^* a_{m-}^* a_{q-} a_{p-} - a_{p+}^* a_{q+}^* a_{m+} a_{0-} \right. \\
 & \quad \left. - a_{0-}^* a_{m+}^* a_{q-} a_{p+} - a_{p-}^* a_{q+}^* a_{m-} a_{0+} - a_{0+}^* a_{m-}^* a_{q+} a_{p-} \right) \\
 & = \sum_{\alpha} \sum_{q \neq 0} J_{q0}^{\alpha} (a_{q+}^* a_{0+} - a_{q-}^* a_{0-}) \sum_{pm \neq 0} J_{pm}^{\alpha} (a_{p+}^* a_{m+} - a_{p-}^* a_{m-}) \\
 & \quad + \sum_{\alpha} \sum_{pm \neq 0} J_{pm}^{\alpha} (a_{p+}^* a_{m+} - a_{p-}^* a_{m-}) \sum_{q \neq 0} J_{q0}^{\alpha} (a_{q+}^* a_{0+} - a_{q-}^* a_{0-})^* \\
 & \geq -\varepsilon'^{-1} \sum_{qq' \neq 0} \widehat{w}_{q0,0q'} a_{q+}^* a_{q'+} a_{0+} a_{0+}^* - \varepsilon'^{-1} \sum_{qq' \neq 0} \widehat{w}_{q0,0q'} a_{q-}^* a_{q'-} a_{0-} a_{0-}^* \\
 & \quad - 2\varepsilon' \sum_{\alpha} \left(\sum_{pm \neq 0} J_{pm}^{\alpha} (a_{p+}^* a_{m+} - a_{p-}^* a_{m-}) \right)^2.
 \end{aligned}$$

The lemma follows from (32). $\square$

Note that if $\varepsilon' \leq 1/4$ then the last term in the estimate in Lemma 7.6 can be controlled by the positive term in Lemma 7.4.

8. Analyzing the Quadratic Hamiltonian

In this section we consider the main part of the Hamiltonian $\widetilde{H}_{r,R}^{\delta}$, namely the “quadratic” Hamiltonian considered by Foldy. As in the previous section we have fixed $z = 0$, i.e., we consider the cube $[-\ell/2, \ell/2]^3$ and we omit z . The “quadratic” Hamiltonian consists of the kinetic energy $\mathcal{K}$ from (13) and all the two-body terms with the coefficients $\widehat{w}_{pq,00}$, $\widehat{w}_{00,pq}$, $\widehat{w}_{p0,0q}$, and $\widehat{w}_{0p,q0}$ with $p, q \neq 0$, i.e.,

$$\begin{aligned}
 H_{\text{Foldy}} &= \sum_{i=1}^N \frac{\gamma_{\varepsilon,i}}{2} \mathcal{K}_i \\
 &+ \frac{1}{2} \sum_{\substack{pq \neq 0 \\ \varepsilon, \varepsilon' = \pm}} \widehat{w}_{pq,00} \varepsilon \varepsilon' \left(2a_{pe}^* a_{0\varepsilon'}^* a_{qe'} a_{0\varepsilon} + a_{pe}^* a_{qe'}^* a_{0\varepsilon'} a_{0\varepsilon} + a_{0\varepsilon}^* a_{0\varepsilon'}^* a_{pe'} a_{qe} \right).
 \end{aligned} \tag{35}$$

In order to compute all the bounds we found it necessary to include the first term in (34) into the “quadratic” Hamiltonian. We therefore define

$$\begin{aligned}
 H_Q &= H_{\text{Foldy}} + (v^+ - v^-) \\
 &\quad \sum_{p \neq 0} \widehat{w}_{p0,00} \left(a_{p+}^* a_{0+} - a_{p-}^* a_{0-} + a_{0+}^* a_{p+} - a_{0-}^* a_{p-} \right).
 \end{aligned} \tag{36}$$

Our goal is to give a lower bound on the ground state energy of the Hamiltonian H_Q .

For any $k \in \mathbb{R}^3$ denote $\chi_k(x) = e^{ikx} \chi(x)$ and define the operators

$$b_{k\pm}^* = (\ell^3 v^{\pm})^{-1/2} a_{\pm}^* (\mathcal{P} \chi_k) a_{0\pm} \quad \text{and} \quad b_{k\pm} = (\ell^3 v^{\pm})^{-1/2} a_{\pm} (\mathcal{P} \chi_k) a_{0\pm}^*, \tag{37}$$

where as before $\mathcal{P}$ denotes the projection orthogonal to constants. (On the subspace where $v^+ = 0$ we set $b_{k+}^* = 0$ and likewise when $v^- = 0$ we set $b_{k-}^* = 0$.) These operators satisfy the commutation relations (note that $a_{0\pm}$ commutes with $a_{\pm}^*(\mathcal{P}\chi_k)$)

$$[b_{k+}, b_{k'-}^*] = [b_{k+}^*, b_{k'-}] = [b_{k-}^*, b_{k'-}] = 0 \quad (38)$$

and¹

$$[b_{k\pm}, b_{k'\pm}^*] = (\ell^3 v^{\pm})^{-1} a_{0\pm}^* a_{0\pm} (\mathcal{P}\chi_k, \mathcal{P}\chi_{k'}) - (\ell^3 v^{\pm})^{-1} a_{\pm}^* (\mathcal{P}\chi_{k'}) a_{\pm} (\mathcal{P}\chi_k).$$

In particular, we get

$$[b_{k\pm}, b_{k\pm}^*] \leq (\ell^3 v^{\pm})^{-1} a_{0\pm}^* a_{0\pm} (\mathcal{P}\chi_k, \mathcal{P}\chi_k) \leq 1. \quad (39)$$

We shall now estimate the “quadratic” Hamiltonian below by an operator that is really quadratic in the operators $b_{k\pm}^*$ and $b_{k\pm}$.

Lemma 8.1 (The quadratic Hamiltonian). *The quadratic Hamiltonian satisfies the lower bound*

$$H_Q \geq \int_{\mathbb{R}^3} h_Q(k) dk - \sum_{pq \neq 0} \hat{w}_{pq,00} (a_{p+}^* a_{q+} + a_{p-}^* a_{q-}), \quad (40)$$

where

$$\begin{aligned} h_Q(k) = & (2\pi)^{-3} \frac{\ell^3 \gamma_{\varepsilon,t}}{4} \frac{|k|^4}{|k|^2 + (\ell t^6)^{-2}} (b_{k+}^* b_{k+} + b_{-k+}^* b_{-k+} + b_{k-}^* b_{k-} + b_{-k-}^* b_{-k-}) \\ & + \frac{\hat{V}_{\varepsilon,R}(k)}{2(2\pi)^3} \left[(v^+ - v^-) \ell^{-3} \left(\widehat{\chi}(k) (v^+)^{1/2} b_{k+}^* + (v^+)^{1/2} b_{-k+} \right. \right. \\ & \quad \left. \left. - (v^-)^{1/2} b_{k-}^* - (v^-)^{1/2} b_{-k-} \right) + \overline{\widehat{\chi}(k)} ((v^+)^{1/2} b_{k+} \right. \\ & \quad \left. + (v^+)^{1/2} b_{-k+}^* - (v^-)^{1/2} b_{k-} - (v^-)^{1/2} b_{-k-}^* \right) \\ & \quad \left. + \sum_{e,e'=\pm} e e' \sqrt{v^e v^{e'}} (b_{ke}^* b_{ke'} + b_{-ke}^* b_{-ke'} + b_{ke}^* b_{-ke'}^* + b_{ke} b_{-ke'}^*) \right]. \end{aligned} \quad (41)$$

Proof. We consider first the kinetic energy

$$\begin{aligned} & \frac{\gamma_{\varepsilon,t}}{2} \sum_{i=1}^N \mathcal{K}_i \\ &= \frac{\gamma_{\varepsilon,t}}{2} \sum_{p,q} (a_{p+}^* a_{q+} + a_{p-}^* a_{q-}) (u_p, \mathcal{K} u_q) \\ &= (2\pi)^{-3} \frac{\gamma_{\varepsilon,t}}{2} \int \frac{|k|^4}{|k|^2 + (\ell t^6)^{-2}} \sum_{p,q \neq 0} (a_{p+}^* a_{q+} + a_{p-}^* a_{q-}) (u_p, \chi_k) (\chi_k, u_q) dk \\ &= (2\pi)^{-3} \frac{\gamma_{\varepsilon,t}}{2} \int \frac{|k|^4}{|k|^2 + (\ell t^6)^{-2}} (a_+^* (\mathcal{P}\chi_k) a_+ (\mathcal{P}\chi_k) + a_-^* (\mathcal{P}\chi_k) a_- (\mathcal{P}\chi_k)) dk. \end{aligned}$$

¹ The corresponding Eq. (25) in [LS0] has a typographical error.

Ground State Energy of the Two-Component Charged Bose Gas

We get the first line in the expression for $h_Q(k)$ if we use that

$$a_{\pm}^*(\mathcal{P}\chi_k)a_{\pm}(\mathcal{P}\chi_k) \geq (v^{\pm})^{-1}a_{\pm}^*(\mathcal{P}\chi_k)a_{0\pm}a_{0\pm}^*a_{\pm}(\mathcal{P}\chi_k) = \ell^3 b_{k\pm}^* b_{k\pm}.$$

Since

$$w_{r,R}(x, y) = (2\pi)^{-3} \int \hat{V}_{r,R}(k) \chi_k(x) \overline{\chi_k(y)} dk,$$

we have that

$$\begin{aligned} \widehat{w}_{pq,00} &= (2\pi\ell)^{-3} \int \hat{V}_{r,R}(k) (u_p, \mathcal{P}\chi_k) (\mathcal{P}\chi_k, u_q) dk \\ &= (2\pi\ell)^{-3} \int \hat{V}_{r,R}(k) (u_p, \mathcal{P}\chi_k) (u_q, \mathcal{P}\chi_{-k}) dk. \end{aligned}$$

If we use that $\hat{V}_{r,R}(k) = \hat{V}_{r,R}(-k)$ and $\overline{\chi(k)} = \chi(-k)$ we arrive at the expression for h_Q . The last sum in (40) comes from commuting $a_{0\pm}^* a_{0\pm}$ to $a_{0\pm} a_{0\pm}^*$. $\square$

Theorem 8.2 (Particular case of Bogolubov's method). *For arbitrary constants $\mathcal{A}, \mathcal{B}_+, \mathcal{B}_- > 0$ and $\kappa \in \mathbb{C}$ we have for each fixed $k \in \mathbb{R}^3$ the following inequality involving the operators $b_{\pm k\pm}$ defined in (37)*

$$\begin{aligned} &\mathcal{A}(b_{k+}^* b_{k+} + b_{-k+}^* b_{-k+} + b_{k-}^* b_{k-} + b_{-k-}^* b_{-k-}) \\ &+ \kappa \left(\mathcal{B}_+^{1/2} b_{k+}^* - \mathcal{B}_-^{1/2} b_{k-}^* + \mathcal{B}_+^{1/2} b_{-k+} - \mathcal{B}_-^{1/2} b_{-k-} \right) \\ &+ \bar{\kappa} \left(\mathcal{B}_+^{1/2} b_{k+} - \mathcal{B}_-^{1/2} b_{k-} + \mathcal{B}_+^{1/2} b_{-k+}^* - \mathcal{B}_-^{1/2} b_{-k-}^* \right) \\ &+ \sum_{e, e' = \pm} \sqrt{\mathcal{B}_e \mathcal{B}_{e'}} \left(b_{ke}^* b_{ke'} + b_{-ke}^* b_{-ke'} + b_{ke}^* b_{-ke'}^* + b_{ke} b_{-ke'} \right) \\ &\geq -(\mathcal{A} + \mathcal{B}_+ + \mathcal{B}_-) + \sqrt{(\mathcal{A} + \mathcal{B}_+ + \mathcal{B}_-)^2 - (\mathcal{B}_+ + \mathcal{B}_-)^2 - |\kappa|^2}. \end{aligned} \quad (42)$$

Proof. Let us introduce

$$\begin{aligned} d_{\pm}^* &= (\mathcal{B}_+ + \mathcal{B}_-)^{-1/2} (\mathcal{B}_+^{1/2} b_{\pm k+}^* - \mathcal{B}_-^{1/2} b_{\pm k-}^*), \\ c_{\pm}^* &= (\mathcal{B}_+ + \mathcal{B}_-)^{-1/2} (\mathcal{B}_-^{1/2} b_{\pm k+}^* + \mathcal{B}_+^{1/2} b_{\pm k-}^*). \end{aligned}$$

These operators are analogous to the c_k and d_k discussed at the end of Sect. 3. According to (38) and (39) these operators satisfy

$$\begin{aligned} [d_+^*, d_-^*] &= 0, \quad [c_+^*, c_-^*] = 0, \quad [d_+, d_+^*] \leq 1, \\ [d_-, d_-^*] &\leq 1, \quad [c_+, c_+^*] \leq 1, \quad [c_-, c_-^*] \leq 1. \end{aligned}$$

The right side of (42) can be rewritten as

$$\begin{aligned} &\mathcal{A}(d_+^* d_+ + d_-^* d_- + c_+^* c_+ + c_-^* c_-) + (\mathcal{B}_+ + \mathcal{B}_-) \\ &\times \left(d_+^* d_+ + d_-^* d_- + d_+^* d_-^* + d_+ d_- + (\mathcal{B}_+ + \mathcal{B}_-)^{-1/2} (\bar{\kappa}(d_+^* + d_-) + \kappa(d_+ + d_-^*)) \right) \end{aligned}$$

We may now complete the squares to write this as

$$\begin{aligned} & \mathcal{A}(c_+^*c_+ + c_-^*c_-) + D(d_+^* + \alpha d_- + a)(d_+^* + \alpha d_- + a)^* \\ & + D(d_-^* + \alpha d_+ + \bar{a})(d_-^* + \alpha d_+ + \bar{a})^* - D\alpha^2([d_+, d_+^*] + [d_-, d_-^*]) - 2D|a|^2 \end{aligned}$$

if

$$D(1 + \alpha^2) = \mathcal{A} + \mathcal{B}_+ + \mathcal{B}_-, \quad 2D\alpha = \mathcal{B}_+ + \mathcal{B}_-, \quad aD(1 + \alpha) = \kappa(\mathcal{B}_+ + \mathcal{B}_-)^{1/2}.$$

We choose the solution $\alpha = 1 + \frac{\mathcal{A}}{\mathcal{B}_+ + \mathcal{B}_-} - \sqrt{\left(1 + \frac{\mathcal{A}}{\mathcal{B}_+ + \mathcal{B}_-}\right)^2 - 1}$. Hence

$$D\alpha^2 = (\mathcal{B}_+ + \mathcal{B}_-)\alpha/2 = \frac{1}{2} \left(\mathcal{A} + \mathcal{B}_+ + \mathcal{B}_- - \sqrt{(\mathcal{A} + \mathcal{B}_+ + \mathcal{B}_-)^2 - (\mathcal{B}_+ + \mathcal{B}_-)^2} \right),$$

$$D|a|^2 = \frac{|\kappa|^2(\mathcal{B}_+ + \mathcal{B}_-)}{D(1 + \alpha^2 + 2\alpha)} = \frac{|\kappa|^2(\mathcal{B}_+ + \mathcal{B}_-)}{\mathcal{A} + 2\mathcal{B}_+ + 2\mathcal{B}_-} \leq \frac{1}{2}|\kappa|^2.$$

□

Usually when applying Bogolubov's method the operators b and b^* satisfy the canonical commutation relations. In this case the lower bound given above is indeed the lowest eigenvalue of the quadratic operator. In our case the operators b and b^* only satisfy the commutation inequalities (39).

We can now apply Bogolubov's method as formulated in the theorem above to the quadratic Hamiltonian H_Q .

Lemma 8.3 (Lower bound on the quadratic Hamiltonian). *The quadratic Hamiltonian satisfies the lower bound*

$$H_Q \geq -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - \frac{1}{2} (v^+ - v^-)^2 \widehat{w}_{00,00} - Ct^{-6} v \ell^{-1}, \quad (43)$$

where I_0 was defined in (5). Likewise,

$$H_{\text{Foldy}} \geq -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - Ct^{-6} v \ell^{-1}. \quad (44)$$

Moreover,

$$H_{\text{Foldy}} \geq \sum_{i=1}^N \frac{\gamma_{\varepsilon,t}}{4} \mathcal{K}_i - Cv^{5/4} \ell^{-3/4} - Ct^{-6} v \ell^{-1}. \quad (45)$$

Proof. We shall use (40). Note first that

$$\sum_{p,q \neq 0} \widehat{w}_{pq,00} (a_{p+}^* a_{q+} + a_{p-}^* a_{q-}) \leq 4\pi \ell^{-3} R^2 (v - \widehat{n}) \leq 4\pi \ell^{-1} v,$$

by (32) and the fact that $R \leq \ell$.

By (41) and Theorem 8.2 with $\mathcal{B}_{\pm} = v^{\pm}$, $\kappa = (v^+ - v^-) \widehat{\chi}(k) \ell^{-3}$, and

$$\mathcal{A} = \left(\frac{\widehat{V}_{r,R}(k)}{2(2\pi)^3} \right)^{-1} (2\pi)^{-3} \frac{\ell^3 \gamma_{\varepsilon,t}}{4} \frac{|k|^4}{|k|^2 + (\ell t^6)^{-2}},$$

Ground State Energy of the Two-Component Charged Bose Gas

we have

$$\begin{aligned} h_Q(k) &\geq \frac{\hat{V}_{r,R}(k)}{2(2\pi)^3} \left(-(\mathcal{A} + v) + \sqrt{(\mathcal{A} + v)^2 - v^2} - (v^+ - v^-)^2 |\widehat{\chi}(k)|^2 \ell^{-6} \right) \\ &= -\frac{1}{2}(2\pi)^{-3} \left(g(k) + \widetilde{f}(k) - \left((g(k) + \widetilde{f}(k))^2 - g(k)^2 \right)^{1/2} \right) \\ &\quad - \frac{\hat{V}_{r,R}(k)}{2(2\pi)^3 \ell^6} (v^+ - v^-)^2 |\widehat{\chi}(k)|^2, \end{aligned}$$

where

$$g(k) = v \hat{V}_{r,R}(k), \quad \widetilde{f}(k) = \frac{\ell^3 \gamma_{\varepsilon,t}}{2} \frac{|k|^4}{|k|^2 + (\ell t^6)^{-2}}.$$

Since

$$g(k) = 4\pi v \left((k^2 + R^2)^{-1} - (k^2 + r^2)^{-1} \right) \leq 4\pi v |k|^{-2}$$

and the expression $g + \widetilde{f} - ((g + \widetilde{f})^2 - g^2)^{1/2}$ is increasing in g for fixed $\widetilde{f}$ we have

$$\begin{aligned} h_Q(k) &\geq -\frac{1}{2}(2\pi)^{-3} \left(4\pi v |k|^{-2} + \widetilde{f}(k) - \left((4\pi v |k|^{-2} + \widetilde{f}(k))^2 - (4\pi v)^2 |k|^{-4} \right)^{1/2} \right) \\ &\quad - \frac{\hat{V}_{r,R}(k)}{2(2\pi)^3 \ell^6} (v^+ - v^-)^2 |\widehat{\chi}(k)|^2. \end{aligned}$$

If we now replace k by $v^{1/4} \ell^{-3/4} \gamma_{\varepsilon,t}^{-1/4} k$, in the k -integration, and observe that

$$\begin{aligned} &\int \frac{\hat{V}_{r,R}(k)}{2(2\pi)^3 \ell^6} |\widehat{\chi}_\ell(k)|^2 dk \\ &= \frac{1}{2} \ell^{-6} \iint \chi_\ell(x) V_{r,R}(x - y) \chi_\ell(y) dx dy = \frac{1}{2} \widehat{w}_{00,00}, \end{aligned}$$

we arrive at

$$H_Q \geq -\gamma_{\varepsilon,t}^{-1/4} I v^{5/4} \ell^{-3/4} - \frac{1}{2} (v^+ - v^-)^2 \widehat{w}_{00,00} - 4\pi v \ell^{-1}$$

where

$$I = \frac{1}{2} (2\pi)^{-3} \int_{\mathbb{R}^3} 4\pi |k|^{-2} + f(k) - \left((4\pi |k|^{-2} + f(k))^2 - (4\pi)^2 |k|^{-4} \right)^{1/2} dk \quad (46)$$

with

$$f(k) = \frac{1}{2} \frac{|k|^4}{|k|^2 + \gamma_{\varepsilon,t}^{1/2} (v \ell)^{-1/2} t^{-12}}.$$

We now study the I -integral in more details. Since the integrand is monotone decreasing in f we get an upper bound to I if we replace f by a lower bound. Let $a =$

$\frac{1}{2}\mathcal{V}_{\varepsilon,t}^{1/2}(\nu\ell)^{-1/2}t^{-12}$. If $|k| \leq (8a)^{1/2}$ we shall simply replace f by 0. If $|k| > (8a)^{1/2}$ we replace f by the lower bound

$$\frac{1}{2}|k|^2 - a.$$

Note that

$$\begin{aligned} & \left((4\pi|k|^{-2} + \frac{1}{2}|k|^2 - a)^2 - (4\pi)^2|k|^{-4} \right)^{1/2} \\ &= \left(\frac{1}{4}|k|^4 + 4\pi + a^2 - (|k|^2 + 8\pi|k|^{-2})a \right)^{1/2} \\ &\geq \left(\left(\frac{1}{4}|k|^4 + 4\pi \right) \left(1 - 4|k|^{-2}a \right) + a^2 \right)^{1/2} \\ &\geq \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} \left(1 - 2|k|^{-2}a + \frac{a^2}{2(|k|^4/4 + 4\pi)} \right. \\ &\quad \left. - \frac{1}{8} \left(4|k|^{-2}a - \frac{a^2}{(|k|^4/4 + 4\pi)} \right)^2 \right). \end{aligned}$$

If $|k| > (8a)^{1/2}$ then $4|k|^{-2}a \geq a^2(|k|^4/4 + 4\pi)^{-1}$ and hence in this case

$$\begin{aligned} & \left((4\pi|k|^{-2} + \frac{1}{2}|k|^2 - a)^2 - (4\pi)^2|k|^{-4} \right)^{1/2} \\ &\geq \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} - 2|k|^{-2}a \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} \\ &\quad + \frac{a^2}{2(|k|^4/4 + 4\pi)^{1/2}} - 2|k|^{-4}a^2 \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} \\ &\geq \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} - a - Ca|k|^{-4} - C|k|^{-6}a^2 \\ &\geq \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} - a - Ca|k|^{-4}. \end{aligned}$$

Thus

$$\begin{aligned} I &\leq \frac{1}{2}(2\pi)^{-3} \left(\int_{|k| < (8a)^{1/2}} 4\pi|k|^{-2} dk + \int_{|k| > (8a)^{1/2}} 4\pi|k|^{-2} + \frac{1}{2}|k|^2 \right. \\ &\quad \left. - \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} + C|k|^{-4}a dk \right) \\ &\leq \frac{1}{2}(2\pi)^{-3} \int_{\mathbb{R}^3} 4\pi|k|^{-2} + \frac{1}{2}|k|^2 - \left(\frac{1}{4}|k|^4 + 4\pi \right)^{1/2} dk + Ca^{1/2} \\ &= I_0 + C(\nu\ell)^{-1/4}t^{-6} \leq I_0 + Cv^{-5/4}\ell^{3/4}t^{-6}\nu\ell^{-1}. \end{aligned}$$

This gives the stated lower bound on H_Q . The proof of (44) is similar except that one should simply use Theorem 8.2 with $\kappa = 0$. The proof of (45) is the same as the proof of (44) except that we, when proving the lower bound on H_{Foldy} , replace $\mathcal{K}$ by $\mathcal{K}/2$ and simply keep the other half $\mathcal{K}/2$ in the lower bound. Note that this will of course change the constants in the lower bound. $\square$

9. A-priori Bounds on the Kinetic Energy, Non-Neutrality, and Excitations

In this section we shall estimate the kinetic energy $\langle \sum_{i=1}^N \mathcal{K}_i \rangle$, the non-neutrality $\langle (\widehat{n}^+ - \widehat{n}^-)^2 \rangle$, and the excitations $\langle \nu - \widehat{n} \rangle$ in a state for which the energy expectation $\langle \widetilde{H} \rangle$ is low. (Recall that $\widetilde{H}$ was defined in (22) and that we are omitting the subscript z .) More precisely, we shall introduce cutoffs

$$R = \omega(t)^{-1} \ell \quad \text{and} \quad r = \ell^{3/2} (\nu \ell)^{-1/2}, \quad (47)$$

and consider a state such that $\langle \widetilde{H}_{r,R}^{\varepsilon/4} \rangle \leq 0$. For technical reasons this assumption will be more appropriate later (see Lemma 10.1). The Lemmas 6.3, 7.1, 7.2, 7.4, 7.5 (Eq. (33)), and 7.6 together with (45) in Lemma 8.3 control all terms in the Hamiltonian $\widetilde{H}_{r,R}^{\varepsilon/4}$ from (27). If we combine all these estimates and assume that we have further cutoffs $0 < r \leq r' \leq R' \leq R$ we obtain the following bound

$$\begin{aligned} \widetilde{H}_{r,R}^{\varepsilon/4} \geq & \sum_{i=1}^N -\frac{\gamma_{\varepsilon,t}}{4} \varepsilon \Delta_{i,\text{Neu}} + \sum_{i=1}^N \frac{\gamma_{\varepsilon,t}}{4} \mathcal{K}_i - C \nu^{5/4} \ell^{-3/4} - C t^{-6} \nu \ell^{-1} \\ & - \frac{1}{2} \nu R'^{-1} - C \nu^+ \nu^- (\varepsilon^{-3/2} r'^{1/2} + \ell^{-3} r'^{2}) \\ & + \frac{1}{2} \widehat{w}'_{00,00} \left[(\widehat{n}^+ - \widehat{n}^-)^2 - \widehat{n}^+ - \widehat{n}^- \right] - 4\pi R'^2 \ell^{-3} |\widehat{n}^+ - \widehat{n}^-| (\nu - \widehat{n}) \\ & + \frac{1}{2} \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2 - \frac{1}{2} r'^{-1} (\nu - \widehat{n}) \\ & - 2\varepsilon' \widehat{w}'_{00,00} \left((\widehat{n}^+ - \widehat{n}^-)^2 + 1 \right) - \varepsilon'^{-1} 8\pi \ell^{-3} R'^2 (\widehat{n} + 1) (\nu - \widehat{n}) \\ & - 2\varepsilon' \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2. \end{aligned}$$

The prime on $\widehat{w}'_{00,00}$ and J' refers to the fact that these quantities are calculated with the r' and R' cutoffs. If we choose $\varepsilon' = 1/8$ and use $\widehat{w}'_{00,00} \leq 4\pi R'^2 \ell^{-3}$ we arrive at

$$\begin{aligned} \widetilde{H}_{r,R}^{\varepsilon/4} \geq & \sum_{i=1}^N -\frac{\gamma_{\varepsilon,t}}{4} \varepsilon \Delta_{i,\text{Neu}} + \sum_{i=1}^N \frac{\gamma_{\varepsilon,t}}{4} \mathcal{K}_i - C \nu^{5/4} \ell^{-3/4} - C t^{-6} \nu \ell^{-1} \\ & - \frac{1}{2} \nu R'^{-1} - C \nu^2 (\varepsilon^{-3/2} r'^{1/2} + \ell^{-3} r'^{2}) - \frac{1}{2} r'^{-1} (\nu - \widehat{n}) \\ & - C \ell^{-3} R'^2 (\widehat{n} + 1) (\nu - \widehat{n} + 1) + \frac{1}{4} \widehat{w}'_{00,00} (\widehat{n}^+ - \widehat{n}^-)^2. \end{aligned} \quad (48)$$

It turns out that the quantity $\nu \ell$ is important and that the main contribution to our energy estimate will come from boxes where $\nu \ell$ is large (but not too large). We shall need the control on the kinetic energy, non-neutrality, and excitations only for boxes where $\nu \ell \geq \varepsilon \omega(t)^2$.

Lemma 9.1 (Bounds on non-neutrality and excitations). *Let R and r be as in (47). There is a constant $C_1 > 0$ such that if $C_1 \varepsilon \omega(t)^2 \leq \nu \ell$ and $C_1 N \ell^3 \leq \varepsilon^3$ then for any state such that $\langle \widetilde{H}_{r,R}^{\varepsilon/4} \rangle \leq 0$ we have*

$$\left\langle \sum_{i=1}^N \mathcal{K}_i \right\rangle \leq C \varepsilon^{-1/2} t^{-2} v^{5/4} \ell^{-3/4} (v\ell)^{1/4} \quad (49)$$

$$\langle v - \widehat{n} \rangle \leq C \varepsilon^{-3/2} t^{-2} (v\ell)^{3/2} \quad (50)$$

and

$$\langle (\widehat{n}^+ - \widehat{n}^-)^2 \rangle \leq C \varepsilon^{-3/2} t^{-2} v (v\ell)^{3/2}. \quad (51)$$

Proof. We introduce the cutoffs

$$R' = a \varepsilon^{1/2} (v\ell)^{-1/2} \ell \quad \text{and} \quad r' = \varepsilon^{-1/4} \ell^{3/2} (v\ell)^{-1/4},$$

where $0 < a$ shall be specified below. We observe first that $r \leq r' \leq R'/2 \leq R/2$, which will, in particular, imply that $\widehat{w}'_{00,00} \geq C^{-1} R^2 \ell^{-3}$. That $r \leq r'$ follows from the assumption that $v\ell \geq C_1 \varepsilon \omega(t)^2 \geq C_1 \varepsilon$ if we choose $C_1 \geq 1$. That $R' \leq R$ follows from the assumption $C_1 \varepsilon \omega(t)^2 \leq v\ell$ if we choose $C_1 \geq a^2$. Finally, $r' \leq R'/2$ follows from the assumption that $C_1 N \ell^3 \leq \varepsilon^3$ if C_1 is chosen large enough depending on a . In fact, we get that $R'/r' = a \varepsilon^{3/4} (v\ell)^{-1/4} \ell^{-1/2} \geq a \varepsilon^{3/4} N^{-1/4} \ell^{-3/4}$ since $v \leq N$.

We see that (30) implies

$$\sum_{i=1}^N -\frac{\gamma_{\varepsilon,t}}{4} \varepsilon \Delta_{i,\text{Neu}} \geq \frac{1}{4} \gamma_{\varepsilon,t} \varepsilon \pi^2 \ell^{-2} (v - \widehat{n}).$$

From the estimate (48) we hence obtain

$$\begin{aligned} \widetilde{H}_{r,R}^{\varepsilon/4} &\geq \sum_{i=1}^N \frac{\gamma_{\varepsilon,t}}{4} \mathcal{K}_i + \left(\frac{1}{4} \gamma_{\varepsilon,t} \pi^2 - \varepsilon^{-3/4} \ell^{1/2} (v\ell)^{1/4} - a^2 C \right) \varepsilon \ell^{-2} (v - \widehat{n}) \\ &\quad + \frac{1}{4} \widehat{w}'_{00,00} (\widehat{n}^+ - \widehat{n}^-)^2 - C v^{5/4} \ell^{-3/4} \left(1 + a^{-1} \varepsilon^{-1/2} (v\ell)^{1/4} + \varepsilon^{-1/2} (v\ell)^{1/4} \right. \\ &\quad \left. (1 + \varepsilon^{-9/8} \ell^{3/4} (v\ell)^{3/8}) + a^2 \varepsilon (v\ell)^{-5/4} \right) - C t^{-6} v \ell^{-1}. \end{aligned}$$

We have here used that $\widehat{n} \leq v$ and that we may estimate $v + 1 \leq 2v$ since we assume that $v \geq 1$. Since $\varepsilon^{-3} \ell^2 (v\ell) \leq \varepsilon^{-3} N \ell^3 \leq C_1^{-1}$ we may rewrite this estimate as

$$\begin{aligned} \widetilde{H}_{r,R}^{\varepsilon/4} &\geq \sum_{i=1}^N \frac{\gamma_{\varepsilon,t}}{4} \mathcal{K}_i + \left(\frac{1}{4} \gamma_{\varepsilon,t} \pi^2 - C_1^{-1/4} - a^2 C \right) \varepsilon \ell^{-2} (v - \widehat{n}) + \frac{1}{4} \widehat{w}'_{00,00} (\widehat{n}^+ - \widehat{n}^-)^2 \\ &\quad - C v^{5/4} \ell^{-3/4} \left(1 + \varepsilon^{-1/2} (v\ell)^{1/4} (1 + C_1^{-3/8} + a^{-1}) + a^2 \varepsilon (v\ell)^{-5/4} \right) \\ &\quad - C (t^{-14} v \ell^{-1} + t^{-22} \ell^{-2}). \end{aligned}$$

If we choose a and C_1 appropriately we see that the second term on the right side above is bounded below by $C^{-1} \varepsilon \ell^{-2} (v - \widehat{n})$. From this and the assumption $C_1 \varepsilon \leq C \varepsilon t^{-8} \leq v\ell$ (recall that $\omega(t) = C t^{-4}$) we easily obtain (49) and (50). We obtain (51) if we use that $\widehat{w}'_{00,00} \geq C^{-1} R^2 \ell^{-3} = C^{-1} \varepsilon (v\ell)^{-1} \ell^{-1}$. $\square$

10. Localization of the Number of Excited Particles

Although the bound on the expectation value $\langle \nu - \hat{n} \rangle$ given in Lemma 9.1 is sufficient for our purposes we still need to know that $\langle (\nu - \hat{n})^2 \rangle \sim \langle \nu - \hat{n} \rangle^2$. We shall however not prove this for a general state with negative energy. Instead we shall show that we may change the ground state, without changing its energy expectation significantly, in such a way that the only occurring values of $\nu - \hat{n}$ are bounded by the quantity on the right side of (50).

To do this we shall use the method of localizing large matrices in Lemma A.1 of Appendix 13. Let Ψ be any normalized wave function with ν particles in the box $\{z\ell\} + [-\ell/2, \ell/2]^3$. We may write $\Psi = \sum_{m=0}^{\nu} c_m \Psi_m$, where for all $m = 0, 1, \dots, \nu$, Ψ_m is a normalized eigenfunction of $\hat{n}$ with eigenvalue m . We may now consider the $(\nu + 1) \times (\nu + 1)$ Hermitian matrix $\mathcal{A}$ with matrix elements $\mathcal{A}_{mm'} = \left(\Psi_m, \tilde{H}_{r,R}^{\varepsilon/4} \Psi_{m'} \right)$.

We shall use Lemma A.1 for this matrix and the vector $\psi = (c_0, \dots, c_\nu)$. We shall choose M in Lemma A.1 to be given by

$$M = \text{Integer part of } \varepsilon^{-3/2} t^{-2} (\nu \ell)^{3/2}. \quad (52)$$

With the assumption in Lemma 9.1 that $\nu \ell \geq C_1 \varepsilon \omega(t)^2$ we may assume (C_1 so large) that $M > 2$. In particular, we may assume that $M \geq C^{-1} \varepsilon^{-3/2} t^{-2} (\nu \ell)^{3/2}$. With the notation in Lemma A.1 we have $\lambda = (\psi, \mathcal{A}\psi) = (\Psi, \tilde{H}_{r,R}^{\varepsilon/4} \Psi)$. Note also that because of the structure of $\tilde{H}_{r,R}^{\varepsilon/4}$ we have, again with the notation in Lemma A.1, that $d_k = 0$ if $k \geq 3$. Since $M > 2$ this means that the second sum in (71) vanishes. We conclude from Lemma A.1 that there exists a normalized wave function $\tilde{\Psi}$ with the property that the only occurring values of $\nu - \hat{n}$ belong to an interval of length M and such that

$$\left(\Psi, \tilde{H}_{r,R}^{\varepsilon/4} \Psi \right) \geq \left(\tilde{\Psi}, \tilde{H}_{r,R}^{\varepsilon/4} \tilde{\Psi} \right) - CM^{-2}(|d_1(\Psi)| + |d_2(\Psi)|).$$

We shall discuss $d_1 = d_1(\Psi)$ and $d_2 = d_2(\Psi)$ in detail below, but first we give the result on the localization of $\hat{n}$ that we shall use.

Lemma 10.1 (Localization of $\hat{n}$). *Let R and r be as in (47) and M as in (52). Let $C_1 > 0$ be the constant in Lemma 9.1 and assume that $C_1 \varepsilon \leq C_1 \varepsilon \omega(t)^2 \leq \nu \ell$ and $C_1 N \ell^3 \leq \varepsilon^3$. Then for any normalized wave function Ψ such that*

$$\left(\Psi, \tilde{H}_{r,R}^{\varepsilon/4} \Psi \right) \leq -M^{-2}(|d_1(\Psi)| + |d_2(\Psi)|) \quad (53)$$

there exists a normalized wave function $\tilde{\Psi}$, which is a linear combination of eigenfunctions of $\nu - \hat{n}$ with eigenvalues less than CM only, such that

$$\left(\Psi, \tilde{H}_{r,R}^{\varepsilon/4} \Psi \right) \geq \left(\tilde{\Psi}, \tilde{H}_{r,R}^{\varepsilon/4} \tilde{\Psi} \right) - M^{-2}(|d_1(\Psi)| + |d_2(\Psi)|). \quad (54)$$

Here $d_1(\Psi)$ and $d_2(\Psi)$ are given as explained in Lemma A.1.

Proof. We choose $\tilde{\Psi}$ as explained above. Then (54) holds. We also know that the possible $\hat{n}$ values of $\tilde{\Psi}$ range in an interval of length M . We do not know however, where this interval is located. The assumption (53) will allow us to say more about the location of the interval.

In fact, it follows from (53) and (54) that $\left(\tilde{\Psi}, \tilde{H}_{r,R}^{\varepsilon/4} \tilde{\Psi} \right) \leq 0$. It is then a consequence of Lemma 9.1 that $\left(\tilde{\Psi}, (\nu - \hat{n}) \tilde{\Psi} \right) \leq CM$. This of course establishes that the allowed values of $\nu - \hat{n}$ are less than $(C + 1)M$ (which we of course just write as CM). $\square$

Our final task in this section is to bound $d_1(\Psi)$ and $d_2(\Psi)$. We have that $d_1(\Psi) = (\Psi, \tilde{H}_{r,R}^{\varepsilon/4}(1)\Psi)$, where $\tilde{H}_{r,R}^{\varepsilon/4}(1)$ is the part of the Hamiltonian $\tilde{H}_{r,R}^{\varepsilon/4}$ containing all the terms with the coefficients $\hat{w}_{pq,\mu\nu}$ for which precisely one or three indices are 0. These are the terms bounded in Lemmas 7.5 and 7.6. These lemmas are stated as one-sided bounds. It is clear from the proof that they could have been stated as two sided bounds. Alternatively we may observe that $\tilde{H}_{r,R}^{\varepsilon/4}(1)$ is transformed to $-\tilde{H}_{r,R}^{\varepsilon/4}(1)$ by the unitary transform which maps all operators $a_{p\pm}^*$ and $a_{p\pm}$ with $p \neq 0$ to $-a_{p\pm}^*$ and $-a_{p\pm}$. This unitary transform leaves the estimates in Lemmas 7.5 and 7.6 invariant. We therefore immediately get the following bound on $d_1(\Psi)$.

Lemma 10.2 (Control of $d_1(\Psi)$). *With the assumptions in Lemma 10.1 we have for all $\varepsilon' > 0$,*

$$|d_1(\Psi)| \leq \varepsilon'^{-1} 8\pi \ell^{-3} R^2(\Psi, (\hat{n} + 1)(v - \hat{n})\Psi) + 2\varepsilon' \hat{w}_{00,00} \left(\Psi, \left((\hat{n}^+ - \hat{n}^-)^2 + 1 \right) \Psi \right) \\ + 2\varepsilon' \sum_{\alpha} \left(\Psi, \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2 \Psi \right).$$

Likewise, we have that $d_2(\Psi) = (\Psi, \tilde{H}_{r,R}^{\varepsilon/4}(2)\Psi)$, where $\tilde{H}_{r,R}^{\varepsilon/4}(2)$ is the part of the Hamiltonian $\tilde{H}_{r,R}^{\varepsilon/4}$ containing all the terms with precisely two $a_{0\pm}$ or two $a_{0\pm}^*$, i.e., these are the terms in the Foldy Hamiltonian, which do not commute with $\hat{n}$.

Lemma 10.3 (Control of d_2). *With the assumptions in Lemma 10.1 there exists a constant $C > 0$ such that*

$$|d_2(\Psi)| \leq C \varepsilon^{-1/2} t^{-2} v^{5/4} \ell^{-3/4} (v\ell)^{1/4} + 8\pi \ell^{-3} R^2(\Psi, (v - \hat{n})(\hat{n} + 1)\Psi).$$

Proof. We consider the unitary transform that replaces all the operators $a_{p\pm}^*$ and $a_{p\pm}$ with $p \neq 0$ by $-ia_{p\pm}^*$ and $ia_{p\pm}$ respectively. Under this transformation the Foldy Hamiltonian (35) changes into an operator that differs from the Foldy Hamiltonian only by a change of sign on the part that we denoted $\tilde{H}_{r,R}^{\varepsilon/4}(2)$. Since both operators satisfy the bound in (44) we conclude that

$$|d_2(\Psi)| \leq \sum_{i=1}^N \frac{\gamma_{\varepsilon,i}}{2} (\Psi, \mathcal{K}_i \Psi) + \frac{1}{2} \sum_{pq \neq 0} \hat{w}_{pq,00} \left(\Psi, \left(2a_{p+}^* a_{0+}^* a_{0+} a_{q+} \right. \right. \\ \left. \left. + 2a_{p-}^* a_{0-}^* a_{0-} a_{q-} - 2a_{p+}^* a_{0-}^* a_{q-} a_{0+} - 2a_{0+}^* a_{q-}^* a_{p+} a_{0-} \right) \Psi \right) \\ + \gamma_{\varepsilon,i}^{-1/4} I_0 v^{5/4} \ell^{-3/4} + C t^{-6} v \ell^{-1}.$$

Ground State Energy of the Two-Component Charged Bose Gas

If we use the representation (31) we find using a Cauchy-Schwarz inequality that

$$\begin{aligned}
 & - \sum_{pq \neq 0} \widehat{w}_{pq,00} (a_{p+}^* a_{q-}^* a_{q-} a_{0+} + a_{0+}^* a_{q-}^* a_{p+} a_{0-}) \\
 & = - \sum_{\alpha} \left(\sum_{p \neq 0} J_{p0}^{\alpha} a_{p+}^* a_{0+} \left(\sum_{q \neq 0} J_{q0}^{\alpha} a_{q-}^* a_{0-} \right)^* + \sum_{q \neq 0} J_{q0}^{\alpha} a_{q-}^* a_{0-} \left(\sum_{p \neq 0} J_{p0}^{\alpha} a_{p+}^* a_{0+} \right)^* \right) \\
 & \leq \sum_{\alpha} \left(\sum_{p \neq 0} J_{p0}^{\alpha} a_{p+}^* a_{0+} \left(\sum_{q \neq 0} J_{q0}^{\alpha} a_{q+}^* a_{0+} \right)^* + \sum_{q \neq 0} J_{q0}^{\alpha} a_{q-}^* a_{0-} \left(\sum_{p \neq 0} J_{p0}^{\alpha} a_{p-}^* a_{0-} \right)^* \right) \\
 & = \sum_{pq \neq 0} \widehat{w}_{pq,00} \left(a_{p+}^* (a_{0+}^* a_{0+} + 1) a_{q+} + a_{p-}^* (a_{0-}^* a_{0-} + 1) a_{q-} \right).
 \end{aligned}$$

The lemma now follows from (32) and from Lemma 9.1. $\square$

11. Lower Bound on the Energy in a Small Cube

Our goal in this section is to prove a lower bound on the expectation $\langle \widetilde{H} \rangle = (\Psi, \widetilde{H} \Psi)$ of the operator $\widetilde{H}$ from (22). We are aiming at the following result.

Theorem 11.1 (Lower bound on $\widetilde{H}$). *There exists a constant $C_0 > 0$ such that for all $0 < \varepsilon < 1$ and $0 < t < 1/2$ with $C_0 N \ell^3 \leq \varepsilon^3$, and $\varepsilon^{-1} t^{-4} \ell < C_0^{-1}$ we have the estimate*

$$\begin{aligned}
 \langle \widetilde{H} \rangle & \geq -\gamma_{\varepsilon,t}^{-1/4} I_0 \langle \widehat{n} \rangle^{5/4} \ell^{-3/4} \\
 & \quad - C(t^{-6} \nu \ell^{-1} + \varepsilon \ell^{-2} + \varepsilon^{-1} t^{-8} \nu) - K_1(\varepsilon, t, N, \ell) \langle \widehat{n} \rangle^{5/4} \ell^{-3/4} \\
 & \quad - \langle \widehat{n} \rangle^{5/3} \ell^{-1/3} (N^{2/5} \ell)^{-5/3} \left(C \varepsilon^{1/6} + K_2(\varepsilon, t, N, \ell) \right), \tag{55}
 \end{aligned}$$

where K_1 and K_2 are sums of a finite number of terms of the form $C \varepsilon^{-a} t^{-b} (N^{2/5} \ell)^c N^{-d}$ with $d > 0$.

Note that the estimate is not formulated as an operator inequality. The lower bound is not a constant nor an expectation value, but a non-linear expression in $\langle \widehat{n} \rangle$. The reason that we give the bound in terms of an expression in $\langle \widehat{n} \rangle$ rather than in terms of ν is that we shall later take into account the term $T(S_0)$, i.e., the kinetic energy between boxes, which depends on $\langle \widehat{n} \rangle$ (see (18)).

In order to arrive at an estimate expressed in $\langle \widehat{n} \rangle$ we shall make repeated use of the following result.

Lemma 11.2. *If $\langle \widetilde{H} \rangle \leq 0$ we have the estimate*

$$\langle n \rangle \geq \nu (1 - C \varepsilon^{-1} (N^{2/5} \ell) N^{-1/15}).$$

In particular, we may assume that the constant C_1 from Lemma 9.1 is so large that if $C_1 N \ell^3 \leq \varepsilon^3$ (i.e., $(N^{2/5} \ell) \leq C_1^{-1} \varepsilon N^{1/15}$) then

$$\nu \leq C \langle \widehat{n} \rangle \quad \text{and} \quad \nu^{5/4} \leq \langle \widehat{n} \rangle^{5/4} (1 + C \varepsilon^{-1} (N^{2/5} \ell) N^{-1/15}).$$

Proof. This follows from Corollary 6.4 since $\varepsilon^{-1}\ell v^{1/3} \leq \varepsilon^{-1}(N^{2/5}\ell)N^{-1/15}$, where we have used that $v \leq N$. $\square$

We shall prove Theorem 11.1 by dividing into several cases. We first prove an a-priori bound on $\tilde{H}$.

Lemma 11.3 (A-priori bound on $\tilde{H}$). *We may assume that the constant C_1 from Lemma 9.1 is large enough so that if $C_1 N \ell^3 \leq \varepsilon^3$ and $\varepsilon^{-1}t^{-4}\ell < C_1^{-1}$ there exists a constant $C > 0$ such that*

$$\begin{aligned} \tilde{H} \geq & -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - C v^{5/4} \ell^{-3/4} \left(\varepsilon^{-1/2} (v\ell)^{1/4} + \varepsilon^{-2} (v\ell)^{3/4} \ell \right) \\ & - C (t^{-6} v \ell^{-1} + \varepsilon \ell^{-2}). \end{aligned} \quad (56)$$

Proof. If $v = 0$ there is nothing to prove. We may therefore assume that $v \geq 1$. We now make the following choices for the cutoffs:

$$r = b\varepsilon^{-1}\ell^2 \quad \text{and} \quad R = \min\{a\varepsilon^{1/2}(v\ell)^{-1/2}, \omega(t)^{-1}\}\ell,$$

where $0 < a, b$ shall be specified below. We first note that $r \leq R$ if C_1 is sufficiently large (depending on a and b). Indeed, either $R/r = (a/b)(\varepsilon^3 v^{-1} \ell^{-3})^{1/2} \geq (a/b)C_1^{1/2}$ (since $v \leq N$) or $R/r = b^{-1}\omega(t)^{-1}\varepsilon\ell^{-1} \geq C^{-1}b^{-1}t^4\varepsilon\ell^{-1} \geq b^{-1}C^{-1}C_1$.

We proceed as in the beginning of Sect. 9, but we now use Lemma 6.2 instead of Lemma 6.3 and (44) instead of (45). We then get that

$$\begin{aligned} \tilde{H} \geq & \sum_{i=1}^N -\frac{\gamma_{\varepsilon,t}}{2} (\varepsilon - \delta) \Delta_{i,\text{Neu}} - \gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - C t^{-6} v \ell^{-1} - C v R^{-1} \\ & - C v^2 (\delta^{-3/2} r^{1/2} + \ell^{-3} r^2) + \frac{1}{2} \widehat{w}_{00,00} \left[(\widehat{n}^+ - \widehat{n}^-)^2 - \widehat{n}^+ - \widehat{n}^- \right] \\ & - 4\pi R^2 \ell^{-3} |\widehat{n}^+ - \widehat{n}^-| (v - \widehat{n}) + \frac{1}{2} \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+a_{\mu}+}^* - a_{p-a_{\mu}-}^*) \right)^2 \\ & - \frac{1}{2} r^{-1} (v - \widehat{n}) - 2\varepsilon' \widehat{w}_{00,00} \left((\widehat{n}^+ - \widehat{n}^-)^2 + 1 \right) \\ & - \varepsilon'^{-1} 8\pi \ell^{-3} R^2 (\widehat{n} + 1) (v - \widehat{n}) - 2\varepsilon' \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+a_{\mu}+}^* - a_{p-a_{\mu}-}^*) \right)^2. \end{aligned}$$

If we choose $\varepsilon' = 1/4$, use that $\widehat{w}_{00,00} \leq 4\pi R^2 \ell^{-3}$, and $\widehat{n} \leq v$ we obtain after inserting the choices of r and R that

$$\begin{aligned} \tilde{H} \geq & \sum_{i=1}^N -\frac{\gamma_{\varepsilon,t}}{2} (\varepsilon - \delta) \Delta_{i,\text{Neu}} - \gamma_{\varepsilon,t}^{1/4} I_0 v^{5/4} \ell^{-3/4} - C t^{-6} v \ell^{-1} \\ & - \frac{1}{2} \varepsilon^{-1} a^{-1} v \ell^{-1} (v\ell)^{1/2} - C v^2 \ell (b^{1/2} \varepsilon^{-1/2} \delta^{-3/2} + \varepsilon^{-2} b^2) \\ & - \frac{1}{2} \varepsilon b^{-1} (v - \widehat{n}) \ell^{-2} - a^2 \varepsilon C \ell^{-1} (v\ell)^{-1} v (v - \widehat{n} + 1). \end{aligned}$$

We have here inserted the upper bound $a\varepsilon(v\ell)^{-1/2}\ell$ for R . This is allowed, since the only term which is not monotone increasing in R , i.e., $\frac{1}{2}vR^{-1}$, can in fact, according to (29), be ignored when $R = \omega(t)^{-1}\ell$. We have again replaced $v + 1$ by $2v$.

Ground State Energy of the Two-Component Charged Bose Gas

If we now choose $\delta = \varepsilon/2$, and again use that

$$\sum_{i=1}^N -\frac{\gamma_{\varepsilon,t}}{2}(\varepsilon - \delta)\Delta_{i,\text{Neu}} \geq \frac{1}{4}\gamma_{\varepsilon,t}\varepsilon\pi^2\ell^{-2}(v - \widehat{n}),$$

we arrive at

$$\begin{aligned} \widetilde{H} \geq & \left(\frac{1}{4}\gamma_{\varepsilon,t}\pi^2 - b^{-1}/2 - a^2C \right) \varepsilon\ell^{-2}(v - \widehat{n}) - \gamma_{\varepsilon,t}^{1/4}I_0v^{5/4}\ell^{-3/4} - Ct^{-6}v\ell^{-1} \\ & - Cv^{5/4}\ell^{-3/4} \left(a^{-1}\varepsilon^{-1/2}(v\ell)^{1/4} + \varepsilon^{-2}(b^{1/2} + b^2)(v\ell)^{3/4}\ell + a^2\varepsilon(v\ell)^{-5/4} \right). \end{aligned}$$

We obtain the results of the lemma if we make appropriate choices of a and b . $\square$

We now use this estimate to control boxes with few or with many particles. We shall show that these boxes do not contribute to the leading order estimate of the energy. Thus the specific form of the first term in (56) is not important. We may simply estimate it by $Cv^{5/4}\ell^{-3/4}$.

Lemma 11.4 (Boxes with few particles). *Let C_1 be the constant in Lemma 9.1 and assume that $v\ell \leq C_1\varepsilon\omega(t)^2$, $C_1N\ell^3 \leq \varepsilon^3$, and $\varepsilon^{-1}t^{-4}\ell < C_1^{-1}$. Then there is a constant $C > 0$ such that*

$$\widetilde{H} \geq -Ct^{-14}v\ell^{-1} - C\varepsilon\ell^{-2} - C\varepsilon^{-1}t^{-8}v.$$

Proof. The result follows immediately from Lemma 11.3 if we simply insert the assumed bound on $v\ell$ and use that $\omega(t) = Ct^{-4}$. $\square$

Lemma 11.5 (Boxes with many particles). *Let C_1 be the constant in Lemma 9.1. Assume that $v\ell \geq \varepsilon^{-4}(N^{2/5}\ell)^{10}$, $C_1N\ell^3 \leq \varepsilon^3$, and $\varepsilon^{-1}t^{-4}\ell < C_1^{-1}$. Then*

$$\begin{aligned} \langle \widetilde{H} \rangle \geq & -\gamma_{\varepsilon,t}^{-1/4}I_0\langle \widehat{n} \rangle^{5/4}\ell^{-3/4} - C\langle \widehat{n} \rangle^{5/4}\ell^{-3/4}\varepsilon^{-1}(N^{2/5}\ell)N^{-1/15} \\ & - C\langle \widehat{n} \rangle^{5/3}\ell^{-1/3} \left(\varepsilon^{1/6}(N^{2/5}\ell)^{-5/3} + \varepsilon^{-2}(N^{2/5}\ell)^{4/3}N^{-3/15} \right) \\ & - C(t^{-14}v\ell^{-1} + t^{-22}\ell^{-2}). \end{aligned}$$

Proof. We may of course assume that $\langle \widetilde{H} \rangle \leq 0$ and hence use the estimates in Lemma 11.2. We now use Lemma 11.3 and insert the estimates

$$\begin{aligned} v^{5/4}\ell^{-3/4}(v\ell)^{1/4} &= v^{5/3}\ell^{-1/3}(v\ell)^{-1/6} \leq v^{5/3}\ell^{-1/3}\varepsilon^{2/3}(N^{2/5}\ell)^{-5/3}, \\ v^{5/4}\ell^{-3/4}(v\ell)^{3/4}\ell &= v^{5/3}\ell^{-1/3}v^{1/3}\ell^{4/3} \leq v^{5/3}\ell^{-1/3}(N^{2/5}\ell)^{4/3}N^{-3/15}. \end{aligned}$$

In the first inequality we used the assumption on $v\ell$. In the last inequality we simply used that $v \leq N$. $\square$

We now restrict attention to boxes with $\varepsilon\omega(t)^2 \leq v\ell \leq \varepsilon^{-4}(N^{2/5}\ell)^{10}$. In the next lemma we shall prove the lower bound on $\langle \widetilde{H} \rangle$ under the restrictive assumption given in (57) below. Finally, Theorem 11.1 is proved by considering the alternative case that (57) fails. Let us note that, logically speaking, this could have been done in the reverse order. I.e., we could, instead, have begun with the case that (57) fails.

Lemma 11.6 (Lower bound on $\tilde{H}$: restricted version). *Let R and r be given by (47). Let C_1 be the constant in Lemma 9.1. Assume that $C_1 \varepsilon \omega(t)^2 \leq v\ell \leq \varepsilon^{-4}(N^{2/5}\ell)^{10}$ and $C_1 N \ell^3 \leq \varepsilon^3$. Then there exists a constant $C_2 > 0$ such that if*

$$\ell^{-3} R^2 v \langle v - \hat{n} \rangle \leq C_2^{-1} \left\langle \hat{w}_{00,00} (\hat{n}^+ - \hat{n}^-)^2 + \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^{*} a_{\mu+} - a_{p-}^{*} a_{\mu-}) \right)^2 \right\rangle, \quad (57)$$

we have that

$$\begin{aligned} \langle \tilde{H} \rangle &\geq -\gamma_{\varepsilon,t}^{-1/4} I_0 \langle \hat{n} \rangle^{5/4} \ell^{-3/4} - C t^{-6} v \ell^{-1} \\ &\quad - C \langle \hat{n} \rangle^{5/4} \ell^{-3/4} \left(\varepsilon^{-1} (N^{2/5} \ell) N^{-1/15} + \varepsilon^{-7/2} (N^{2/5} \ell)^{23/4} N^{-3/10} \right. \\ &\quad \left. + \varepsilon^{-9/2} t^{-2} (N^{2/5} \ell)^8 N^{-1/5} \right). \end{aligned}$$

Proof. We may of course assume that $\langle \tilde{H} \rangle \leq 0$ and hence use Lemma 11.2. We may also assume that $\langle \tilde{H}_{r,R}^{5/4} \rangle \leq 0$ otherwise we have from (29) and the choice of r that

$$\begin{aligned} \langle \tilde{H} \rangle &\geq -C v^2 (\varepsilon^{-3/2} r^{1/2} + \ell^{-3} r^2) = -C \varepsilon^{-3/2} v^{5/4} \ell^{-3/4} (v \ell)^{1/2} \ell^{3/4} - C v \ell^{-1} \\ &\geq -C \varepsilon^{-3/2} \langle \hat{n} \rangle^{5/4} \ell^{-3/4} (\varepsilon^{-4} (N^{2/5} \ell)^{10})^{1/2} (N^{2/5} \ell)^{3/4} N^{-3/10} - C v \ell^{-1} \\ &= -C \langle \hat{n} \rangle^{5/4} \varepsilon^{-7/2} \ell^{-3/4} (N^{2/5} \ell)^{23/4} N^{-3/10} - C v \ell^{-1}, \end{aligned}$$

which implies the stated estimate. It now follows from Lemma 9.1 that we have the estimate (50).

We again proceed as in the beginning of Sect. 9, but we use (44) instead of (45) and (29) instead of Lemma 6.3 or (28) (since now $R = \omega(t)^{-1} \ell$) and we choose $\delta = \varepsilon$. We then get that

$$\begin{aligned} \tilde{H} &\geq -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - C t^{-6} v \ell^{-1} - C v^2 (\varepsilon^{-3/2} r^{1/2} + \ell^{-3} r^2) \\ &\quad + \frac{1}{2} \hat{w}_{00,00} \left[(\hat{n}^+ - \hat{n}^-)^2 - \hat{n}^+ - \hat{n}^- \right] - 4\pi R^2 \ell^{-3} |\hat{n}^+ - \hat{n}^-| (v - \hat{n}) \\ &\quad + \frac{1}{2} \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^{*} a_{\mu+} - a_{p-}^{*} a_{\mu-}) \right)^2 \\ &\quad - \frac{1}{2} r^{-1} (v - \hat{n}) - 2\varepsilon' \hat{w}_{00,00} \left((\hat{n}^+ - \hat{n}^-)^2 + 1 \right) \\ &\quad - \varepsilon'^{-1} 8\pi \ell^{-3} R^2 (\hat{n} + 1) (v - \hat{n}) - 2\varepsilon' \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^{*} a_{\mu+} - a_{p-}^{*} a_{\mu-}) \right)^2. \end{aligned}$$

If we use the assumption (57) and the facts that $|\hat{n}^+ - \hat{n}^-| \leq v$, $\hat{n} + 1 \leq 2v$, and $\hat{w}_{00,00} \leq 4\pi R^2 \ell^{-3}$ we see with appropriate choices of ε' and C_2 that

$$\begin{aligned} \langle \tilde{H} \rangle &\geq -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - C (t^{-14} v \ell^{-1} + t^{-22} \ell^{-2}) \\ &\quad - C \varepsilon^{-3/2} v^{7/4} \ell^{1/2} - C v \ell^{-1} - C \omega(t)^{-2} v \ell^{-1} - C (v - \hat{n}) v^{1/2} \ell^{-1}, \end{aligned}$$

where we have inserted the choices of r and R . If we use (50) we see that

$$\begin{aligned} (v - \hat{n}) v^{1/2} \ell^{-1} &\leq C \varepsilon^{-3/2} t^{-2} v^{5/4} \ell^{-3/4} (v \ell)^{3/4} \ell^{1/2} \\ &\leq C \varepsilon^{-9/2} t^{-2} v^{5/4} \ell^{-3/4} (N^{2/5} \ell)^8 N^{-1/5}. \end{aligned}$$

Ground State Energy of the Two-Component Charged Bose Gas

Hence we arrive at the bound in the lemma if we note, as above, that

$$\varepsilon^{-3/2} \nu^{7/4} \ell^{1/2} \leq C \nu^{5/4} \ell^{-3/4} \varepsilon^{-7/2} (N^{2/5} \ell)^{23/4} N^{-3/10}$$

and use the result of Lemma 11.2. $\square$

Proof of Theorem 11.1. We again let r and R be given by (47). According to Lemmas 11.4 and 11.5 we may assume that $C_1 \varepsilon \omega(t)^2 \leq \nu \ell \leq \varepsilon^{-4} (N^{2/5} \ell)^{10}$. Then from Lemma 11.6 we may assume that

$$\begin{aligned} & \ell^{-3} R^2 \nu \langle \nu - \widehat{n} \rangle \\ & \geq C_2^{-1} \left\langle \widehat{w}_{00,00} (\widehat{n}^+ - \widehat{n}^-)^2 + \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2 \right\rangle. \end{aligned} \quad (58)$$

We may also still assume that $\langle \widetilde{H} \rangle \leq 0$ and $\langle \widetilde{H}_{r,R}^{\varepsilon/4} \rangle \leq 0$ (as in the proof of Lemma 11.6). Hence we can use the results of Lemmas 9.1 and 11.2. It is enough to prove an estimate of the type (55) for $\langle \widetilde{H}_{r,R}^{\varepsilon/4} \rangle$ (the extra error terms were estimated in the first paragraph of the proof of Lemma 11.6).

We begin by bounding d_1 and d_2 using Lemmas 10.2 and 10.3. We have from Lemmas 9.1 and 10.3 that

$$\begin{aligned} |d_2| & \leq C \varepsilon^{-1/2} t^{-2} \nu^{5/4} \ell^{-3/4} (\nu \ell)^{1/4} + C \ell^{-3} R^2 \nu \langle \nu - \widehat{n} \rangle \\ & \leq C \nu^{5/4} \ell^{-3/4} \left(\varepsilon^{-1/2} t^{-2} (\nu \ell)^{1/4} + t^6 \varepsilon^{-3/2} (\nu \ell)^{5/4} \right) \leq C t^6 \varepsilon^{-3/2} \nu^{5/4} \ell^{-3/4} (\nu \ell)^{5/4}, \end{aligned}$$

where the last estimate follows since $\nu \ell \geq \varepsilon \omega(t)^2$.

In order to bound d_1 we shall use (58). Together with Lemma 10.2 this gives (choosing $\varepsilon' = 1/2$ say)

$$|d_1| \leq C \ell^{-3} R^2 \nu (\langle \nu - \widehat{n} \rangle + 1) \leq C t^6 \varepsilon^{-3/2} \nu^{5/4} \ell^{-3/4} (\nu \ell)^{5/4}.$$

If the assumption (53) fails then

$$\begin{aligned} \langle \widetilde{H}_{r,R}^{\varepsilon/4} \rangle & \geq -C t^{10} \varepsilon^{3/2} \nu^{5/4} \ell^{-3/4} (\nu \ell)^{-7/4} \\ & \geq -C t^{10} \varepsilon^{3/2} \nu \ell^{-1} (\varepsilon \omega(t)^2)^{-3/2} = -C t^{22} \nu \ell^{-1}, \end{aligned}$$

and we see then that the bound (55) holds. We may therefore assume that (53) holds. Thus from Lemma 10.1 it follows that we can find a normalized wave function $\widetilde{\Psi}$, such that

$$(\widetilde{\Psi}, (\nu - \widehat{n}) \widetilde{\Psi}) \leq C \varepsilon^{-3/2} t^{-2} (\nu \ell)^{3/2} \quad \text{and} \quad (\widetilde{\Psi}, (\nu - \widehat{n})^2 \widetilde{\Psi}) \leq C \varepsilon^{-3} t^{-4} (\nu \ell)^3 \quad (59)$$

and such that

$$\langle \widetilde{H}_{r,R}^{\varepsilon/4} \rangle \geq (\widetilde{\Psi}, \widetilde{H}_{r,R}^{\varepsilon/4} \widetilde{\Psi}) - C t^{22} \nu \ell^{-1}. \quad (60)$$

(Here as before $\langle X \rangle$ denotes the expectation of the operator X in the state Ψ .)

In order to analyze $(\widetilde{\Psi}, \widetilde{H}_{r,R}^{\varepsilon/4} \widetilde{\Psi})$ we again proceed as in the beginning of Sect. 9, but this time we use (34) of Lemma 7.5 and (43) of Lemma 8.3. We find that

$$\begin{aligned}
 \tilde{H}_{r,R}^{\varepsilon/4} &\geq -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - C t^{-6} v \ell^{-1} \\
 &\quad + \frac{1}{2} \widehat{w}_{00,00} \left[(\widehat{n}^+ - \widehat{n}^-)^2 - \widehat{n}^+ - \widehat{n}^- \right] - 4\pi R^2 \ell^{-3} |\widehat{n}^+ - \widehat{n}^-| (v - \widehat{n}) \\
 &\quad - \frac{1}{2} \widehat{w}_{00,00} (v^+ - v^-)^2 + \frac{1}{2} \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2 \\
 &\quad - \frac{1}{2} r^{-1} (v - \widehat{n}) - 2\varepsilon' \widehat{w}_{00,00} \left((v - \widehat{n})^2 + 1 \right) - \varepsilon'^{-1} 8\pi \ell^{-3} R^2 (\widehat{n} + 1) (v - \widehat{n}) \\
 &\quad - 2\varepsilon' \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2. \tag{61}
 \end{aligned}$$

We shall use that for all $\varepsilon'' > 0$

$$(v^+ - v^-)^2 \leq (1 + \frac{\varepsilon''}{2})(\widehat{n}^+ - \widehat{n}^-)^2 + (1 + \frac{1}{2\varepsilon''})(v - \widehat{n})^2$$

and

$$|\widehat{n}^+ - \widehat{n}^-| (v - \widehat{n}) \leq \frac{\varepsilon''}{2} (\widehat{n}^+ - \widehat{n}^-)^2 + \frac{1}{2\varepsilon''} (v - \widehat{n})^2.$$

According to Lemma 7.4

$$\begin{aligned}
 \sum_{\alpha} \left(\sum_{p\mu \neq 0} J_{p\mu}^{\alpha} (a_{p+}^* a_{\mu+} - a_{p-}^* a_{\mu-}) \right)^2 &\leq \sum_{pq, \mu v \neq 0} \widehat{w}_{pq, \mu v} \left(a_{p+}^* a_{q+}^* a_{v+} a_{\mu+} \right. \\
 &\quad \left. + a_{p-}^* a_{q-}^* a_{v-} a_{\mu-} - 2a_{p+}^* a_{q-}^* a_{v-} a_{\mu+} \right) + (v - \widehat{n}) r^{-1}.
 \end{aligned}$$

The sum on the right is simply the second quantization of the two-body operator $\mathcal{P} \otimes \mathcal{P} w_{r,R} \mathcal{P} \otimes \mathcal{P}$, where $\mathcal{P}$ again denotes the projection onto the subspace of $L^2([- \ell/2, \ell/2]^3) \times \{1, -1\}$ consisting of functions orthogonal to constants. Since $w_{r,R} \leq r^{-1}$ this sum is bounded above by

$$\left((v^+ - \widehat{n}^+)^2 + (v^- - \widehat{n}^-)^2 \right) r^{-1} \leq (v - \widehat{n})^2 r^{-1}.$$

(This can of course also be proved by directly estimating the sum.) If we insert the above estimates into (61) and choose $\varepsilon' = 1/\varepsilon''$ we arrive at

$$\begin{aligned}
 \tilde{H}_{r,R}^{\varepsilon/4} &\geq -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - C t^{-6} v \ell^{-1} \\
 &\quad - C R^2 \ell^{-3} \left(v + \varepsilon'' (\widehat{n}^+ - \widehat{n}^-)^2 + (1 + \frac{1}{\varepsilon''}) \left((v - \widehat{n})^2 + 1 \right) + \varepsilon'' v (v - \widehat{n}) \right) \\
 &\quad - C r^{-1} \left(\frac{1}{\varepsilon''} (v - \widehat{n})^2 + (1 + \frac{1}{\varepsilon''}) (v - \widehat{n}) \right).
 \end{aligned}$$

We now use (59), Lemma 9.1 (we may of course assume that $(\tilde{\Psi}, \tilde{H}_{r,R}^{\varepsilon/4} \tilde{\Psi}) \leq 0$), and the choices (47) of R and r . We get

$$\begin{aligned}
 (\tilde{\Psi}, \tilde{H}_{r,R}^{\varepsilon/4} \tilde{\Psi}) &\geq -\gamma_{\varepsilon,t}^{-1/4} I_0 v^{5/4} \ell^{-3/4} - C t^{-6} v \ell^{-1} \\
 &\quad - C \omega(t)^{-2} \ell^{-1} \left(v + \varepsilon'' \varepsilon^{-3/2} t^{-2} v (v \ell)^{3/2} + (1 + \frac{1}{\varepsilon''}) \varepsilon^{-3} t^{-4} (v \ell)^3 \right) \\
 &\quad - C \ell^{-3/2} (v \ell)^{1/2} \left(\frac{1}{\varepsilon''} \varepsilon^{-3} t^{-4} (v \ell)^3 + (1 + \frac{1}{\varepsilon''}) \varepsilon^{-3/2} t^{-2} (v \ell)^{3/2} \right).
 \end{aligned}$$

Ground State Energy of the Two-Component Charged Bose Gas

Finally we make the choice $\varepsilon'' = \varepsilon^{-3/4} t^{-5} (\nu \ell)^{1/2} \ell^{1/4}$ and arrive at

$$\begin{aligned} \left(\tilde{\Psi}, \tilde{H}_{r,R}^{\varepsilon/4} \tilde{\Psi} \right) &\geq -\gamma_{\varepsilon,t}^{-1/4} I_0 \nu^{5/4} \ell^{-3/4} - C t^{-6} \nu \ell^{-1} - C \nu^{5/4} \ell^{-3/4} \\ &\quad \times \left(\varepsilon^{-9/4} t^1 (\nu \ell)^{7/4} \ell^{1/4} + \varepsilon^{-3} t^4 (\nu \ell)^{7/4} \ell + \varepsilon^{9/4} t^9 (\nu \ell)^{5/4} \ell^{3/4} \right. \\ &\quad \left. + \varepsilon^{-3/2} t^{-2} (\nu \ell)^{3/4} \ell^{1/2} + \varepsilon^{-3/4} t^3 (\nu \ell)^{1/4} \ell^{1/4} \right). \end{aligned}$$

If we now insert the bound $\nu \ell \leq \varepsilon^{-4} (N^{2/5} \ell)^{10}$, use (60) and the result of Lemma 11.2 we arrive at a bound of the form (55) for $\langle \tilde{H}_{r,R}^{\varepsilon/4} \rangle$. $\square$

12. The Lattice Approximation

Given a map $S : \mathbb{Z}^3 \rightarrow \mathbb{R}$ we define a function $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ as follows. On any cube $\{\mu\} + [-1/2, 1/2]^3$ with $\mu \in \{(1/2, 1/2, 1/2)\} + \mathbb{Z}^3$ we set

$$\phi(x) = \sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau| = \sqrt{3}}} \lambda_\tau (x - \mu) S(\mu + \tau/2), \quad (62)$$

where

$$\lambda_\tau(x) = (1/2 + \tau_1 x_1)(1/2 + \tau_2 x_2)(1/2 + \tau_3 x_3). \quad (63)$$

Note that the 8 points $\mu + \tau/2$ with $|\tau| = \sqrt{3}$ are the corners of the cube $\{\mu\} + [-1/2, 1/2]^3$. The function ϕ is well-defined on $\mathbb{R}^3$ and is continuous. Moreover, $\phi(\sigma) = S(\sigma)$ for $\sigma \in \mathbb{Z}^3$. By a straightforward calculation we obtain that

$$\int_{\mathbb{R}^3} (\nabla \phi)^2 = T(S), \quad (64)$$

where we have defined

$$T(S) = \sum_{\substack{\sigma_1, \sigma_2 \in \mathbb{Z}^3 \\ |\sigma_1 - \sigma_2| = \sqrt{2}}} \frac{1}{12} (S(\sigma_1) - S(\sigma_2))^2 + \sum_{\substack{\sigma_1, \sigma_2 \in \mathbb{Z}^3 \\ |\sigma_1 - \sigma_2| = \sqrt{3}}} \frac{1}{24} (S(\sigma_1) - S(\sigma_2))^2.$$

Note that there is a constant $C > 0$ such that for all maps $S : \mathbb{Z}^3 \rightarrow \mathbb{R}$ we have

$$C^{-1} \sum_{\substack{\sigma_1, \sigma_2 \in \mathbb{Z}^3 \\ |\sigma_1 - \sigma_2| = 1}} (S(\sigma_1) - S(\sigma_2))^2 \leq T(S) \leq C \sum_{\substack{\sigma_1, \sigma_2 \in \mathbb{Z}^3 \\ |\sigma_1 - \sigma_2| = 1}} (S(\sigma_1) - S(\sigma_2))^2.$$

Lemma 12.1. *If $\lambda_1, \dots, \lambda_m \geq 0$ and $\sum_{j=1}^m \lambda_j = 1$ we have for all $S_1, \dots, S_m \geq 0$ and all $\beta \geq 1$ that*

$$\left(\sum_{j=1}^m \lambda_j S_j \right)^\beta \leq \sum_{j=1}^m \lambda_j S_j^\beta \leq \left(\sum_{j=1}^m \lambda_j S_j \right)^\beta + C_{m,\beta} \left(\sum_{1 \leq i < j \leq m} (S_i - S_j)^2 \right)^{1/2} \sum_{j=1}^m S_j^{\beta-1}$$

for some constant $C_{m,\beta} > 0$.

Proof. The first inequality follows from an application of Jensen's inequality. If we write $Y = \sum_{j=1}^m \lambda_j S_j$ then $\sum_{j=1}^m \lambda_j S_j^\beta - \left(\sum_{j=1}^m \lambda_j S_j \right)^\beta = \sum_{j=1}^m \lambda_j (S_j^\beta - Y^\beta)$. The second inequality above then follows easily from $|a^\beta - b^\beta| \leq \beta|a - b|(a^{\beta-1} + b^{\beta-1})$, which holds for all $a, b > 0$, and $|S_i - Y| \leq \sum_{j=1}^m \lambda_j |S_i - S_j|$. $\square$

Lemma 12.2. *If $\beta \geq 2$ and if $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ is the function constructed above corresponding to a non-negative map $S : \mathbb{Z}^3 \rightarrow \mathbb{R}$ we have for all $0 < \delta$ that*

$$(1 - \delta) \sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^\beta - C\delta^{-(\beta-1)} T(S)^{\beta/2} \leq \int_{\mathbb{R}^3} \phi^\beta \leq \sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^\beta$$

and that

$$\begin{aligned} & \sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^{5/2} - \delta T(S) - C\delta^{-1} \left(\sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^6 \right)^{1/4} \left(\sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^2 \right)^{3/4} \\ & \leq \int_{\mathbb{R}^3} \phi^{5/2} \leq \sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^{5/2} \end{aligned}$$

Proof. The functions $\lambda_\tau(x)$ in (63) are non-negative and satisfy

$$\sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau|=\sqrt{3}}} \lambda_\tau(x) = 1$$

and that $\int_{[-1/2, 1/2]^3} \lambda_\tau(x) dx = 1/8$ for all $\tau \in \mathbb{Z}^3$ with $|\tau| = \sqrt{3}$. We thus get from the inequality in Lemma 12.1 that for all $\mu \in \{(1/2, 1/2, 1/2)\} + \mathbb{Z}^3$,

$$\begin{aligned} & \frac{1}{8} \sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau|=\sqrt{3}}} S(\mu + \tau/2)^\beta \\ & \geq \int_{\{\mu\} + [-1/2, 1/2]^3} \phi(x)^\beta dx \geq \frac{1}{8} \sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau|=\sqrt{3}}} S(\mu + \tau/2)^\beta \\ & \quad - C \left(\sum_{\substack{\tau_1, \tau_2 \in \mathbb{Z}^3 \\ |\tau_1|=\sqrt{3} \\ |\tau_2|=\sqrt{3}}} (S(\mu + \tau_1/2) - S(\mu + \tau_2/2))^2 \right)^{1/2} \sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau|=\sqrt{3}}} S(\mu + \tau/2)^{\beta-1} \\ & \geq \frac{1}{8} (1 - \delta) \sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau|=\sqrt{3}}} S(\mu + \tau/2)^\beta \\ & \quad - C\delta^{-(\beta-1)} \left(\sum_{\substack{\tau_1, \tau_2 \in \mathbb{Z}^3 \\ |\tau_1|=\sqrt{3} \\ |\tau_2|=\sqrt{3}}} (S(\mu + \tau_1/2) - S(\mu + \tau_2/2))^2 \right)^{\beta/2}. \end{aligned}$$

If we sum these inequalities over μ (i.e., over cubes), use that each point in $\mathbb{Z}^3$ is the corner of 8 different cubes, and that $\sum_\mu b_\mu^{\beta/2} \leq (\sum_\mu b_\mu)^{\beta/2}$, when $b_\mu \geq 0$, we obtain the first inequality of the lemma.

Ground State Energy of the Two-Component Charged Bose Gas

To arrive at the second inequality we instead use the Cauchy-Schwarz estimate

$$\begin{aligned} & \left(\sum_{\substack{\tau_1, \tau_2 \in \mathbb{Z}^3 \\ |\tau_1| = \sqrt{3} \\ |\tau_2| = \sqrt{3}}} (S(\mu + \tau_1/2) - S(\mu + \tau_2/2))^2 \right)^{1/2} \sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau| = \sqrt{3}}} S(\mu + \tau/2)^{\beta-1} \\ & \leq \delta \sum_{\substack{\tau_1, \tau_2 \in \mathbb{Z}^3 \\ |\tau_1| = \sqrt{3} \\ |\tau_2| = \sqrt{3}}} (S(\mu + \tau_1/2) - S(\mu + \tau_2/2))^2 + C\delta^{-1} \sum_{\substack{\tau \in \mathbb{Z}^3 \\ |\tau| = \sqrt{3}}} S(\mu + \tau/2)^{2\beta-2}. \end{aligned}$$

For $\beta = 5/2$ we have $2\beta - 2 = 3$. The second inequality in the lemma then follows by summing over μ as above, possibly redefining δ , and using that

$$\sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^3 \leq \left(\sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^6 \right)^{1/4} \left(\sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^2 \right)^{3/4}.$$

□

Theorem 12.3 (The Sobolev inequality for T). *There exists a constant $C > 0$ such that for all maps $S : \mathbb{Z}^3 \rightarrow \mathbb{R}$ that vanish outside some finite set we have*

$$\left(\sum_{\sigma \in \mathbb{Z}^3} |S(\sigma)|^6 \right)^{1/3} \leq CT(S).$$

Proof. Since $T(S) \geq T(|S|)$ it is enough to consider only non-negative maps S . If we use the first inequality in Lemma 12.2 with $\beta = 6$ and $\delta = 1/2$ we obtain that

$$\sum_{\sigma \in \mathbb{Z}^3} |S(\sigma)|^6 \leq 2 \int \phi(x)^6 dx + CT(S)^3.$$

Since $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ is a C^1 function of compact support we may use the standard Sobolev inequality on $\mathbb{R}^3$

$$\int \phi(x)^6 dx \leq C \left(\int |\nabla \phi(x)|^2 dx \right)^3 = CT(S)^3.$$

□

If S vanishes outside a finite set we get from the Sobolev inequality and the second inequality in Lemma 12.2 that

$$\sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^{5/2} - \delta T(S) - C\delta^{-7} \left(\sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^2 \right)^3 \leq \int_{\mathbb{R}^3} \phi^{5/2} \leq \sum_{\sigma \in \mathbb{Z}^3} S(\sigma)^{5/2}. \quad (65)$$

12.1. Application of the lattice approximation to the situation in Theorem 5.2. We shall use the above results for the map

$$S_0(\sigma) = S_0^\psi(\sigma) = \ell^{-1}(\sqrt{\langle n_\sigma \rangle + 1} - 1).$$

The expression $T(S_0)$ is then exactly the term that appears in Theorem 5.2 on localizing into small cubes. Note that $S_0(\sigma) \leq \ell^{-1}\sqrt{\langle n_\sigma \rangle}$. Let $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ be the function corresponding to S_0 as constructed in (62). Then from (64) and Lemma 12.2 we get

$$T(S_0) = \int (\nabla \phi)^2 \quad \text{and} \quad \ell^2 \int \phi^2 \leq \sum_{\sigma \in \mathbb{Z}^3} \ell^2 S_0(\sigma)^2 \leq \sum_{\sigma \in \mathbb{Z}^3} \langle n_\sigma \rangle \leq N. \quad (66)$$

Moreover, we get for all $0 < \delta$ that

$$\begin{aligned} & \sum_{\sigma \in \mathbb{Z}^3} \langle n_\sigma \rangle^{5/4} \ell^{-3/4} \\ & \leq (1 - \delta) \sum_{\sigma \in \mathbb{Z}^3} S_0(\sigma)^{5/2} \ell^{7/4} + C \delta^{-3/2} \ell^{-3/4} (L/\ell)^3 \\ & \leq (1 - \delta) \ell^{7/4} \int \phi^{5/2} + \delta T(S_0) + C \delta^{-7} \ell^8 \left(\sum_{\sigma \in \mathbb{Z}^3} \ell^2 S_0(\sigma)^2 \right)^3 + C \delta^{-3/2} \ell^{-3/4} (L/\ell)^3 \\ & \leq (1 - \delta) \ell^{7/4} \int \phi^{5/2} + \delta T(S_0) + C \delta^{-7} \ell^8 N^3 + C \delta^{-3/2} \ell^{-3/4} (L/\ell)^3, \end{aligned} \quad (67)$$

where we have used firstly that $\langle n_\sigma \rangle \neq 0$ for at most $(L/\ell)^3$ different points σ and secondly the inequality (65) with δ replaced by $\delta \ell^{-7/4}$. Likewise we get from the Sobolev inequality Theorem 12.3

$$\begin{aligned} \sum_{\sigma \in \mathbb{Z}^3} \langle n_\sigma \rangle^{5/3} \ell^{-1/3} & \leq C \sum_{\sigma \in \mathbb{Z}^3} \ell^3 S_0(\sigma)^{10/3} + C \ell^{-1/3} (L/\ell)^3 \\ & \leq C \left(\sum_{\sigma \in \mathbb{Z}^3} S_0(\sigma)^6 \right)^{1/3} \left(\ell^{5/2} \sum_{\sigma \in \mathbb{Z}^3} \ell^2 S_0(\sigma)^2 \right)^{2/3} + C \ell^{-1/3} (L/\ell)^3 \\ & \leq C T(S_0) (N^{2/5} \ell)^{5/3} + C \ell^{-1/3} (L/\ell)^3. \end{aligned} \quad (68)$$

Finally, note that if we define $\tilde{\Phi}(x) = \ell^{-1/2} \phi(x/\ell)$ we find that

$$\int (\nabla \tilde{\Phi})^2 = \int (\nabla \phi)^2, \quad \int \tilde{\Phi}^{5/2} = \ell^{7/4} \int \phi^{5/2}, \quad \int \tilde{\Phi}^2 = \ell^2 \int \phi^2. \quad (69)$$

13. Completing the Proof of the Lower Bound in Dyson's Formula Theorem 2.1

In this final section we shall combine all the previous results to conclude the asymptotic lower bound in Theorem 2.1. From Theorems 4.2, 5.2, 11.1, and (66), (67), (68), and (69) we obtain, under the assumptions that $C_0 N \ell^3 \leq \varepsilon^3$ and $\varepsilon^{-1} \ell^{-4} \leq C_0^{-1}$, that

$$E(N) \geq \inf_{\tilde{\Phi}} \left(A \int (\nabla \tilde{\Phi})^2 - B \int \tilde{\Phi}^{5/2} \right) - D N^{7/5},$$

Ground State Energy of the Two-Component Charged Bose Gas

where the infimum is over all functions $\tilde{\Phi} \geq 0$, such that $\int \tilde{\Phi}^2 \leq N$ and

$$\begin{aligned} A &= (\gamma \tilde{\gamma} \gamma_{\varepsilon,t}/2 - \gamma \tilde{\gamma} \gamma_{\varepsilon,t}^{-1/4} I_0 \delta - C \varepsilon^{1/6} - C K_2(\varepsilon, t, N, \ell)) \\ B &= (1 - \delta) \gamma \tilde{\gamma} (\gamma_{\varepsilon,t}^{-1/4} I_0 + K_1(\varepsilon, t, N, \ell)) \\ D &= C \left(t^{-2} (N^{1/5} L)^{-2} + t^{-4} (N^{2/5} \ell)^{-1} + (N^{1/5} L)^3 (N^{2/5} \ell)^{-5} \right. \\ &\quad + t^{-6} (N^{1/5} \ell)^{-1} N^{-1/5} + \varepsilon^{-1} t^{-8} N^{-2/5} \\ &\quad + \gamma \tilde{\gamma} (\gamma_{\varepsilon,t}^{-1/4} I_0 + K_1(\varepsilon, t, N, \ell)) \delta^{-7} (N^{2/5} \ell)^8 N^{-8/5} \\ &\quad + \gamma \tilde{\gamma} (\gamma_{\varepsilon,t}^{-1/4} I_0 + K_1(\varepsilon, t, N, \ell)) \delta^{-3/2} (N^{1/5} L)^3 (N^{2/5} \ell)^{-15/4} N^{-1/2} \\ &\quad \left. + (\varepsilon^{1/6} + K_2(\varepsilon, t, N, \ell)) (N^{2/5} \ell)^{-5} (N^{1/5} L)^3 N^{-2/3} \right). \end{aligned}$$

We have used that $\sum_{\sigma} v_{\sigma} = N$ and $\sum_{\sigma, v_{\sigma} \neq 0} \ell^{-2} \leq L^3 / \ell^5$.

We shall show that as $N \rightarrow \infty$ we can choose $\varepsilon, t, \delta \rightarrow 0$ and $N^{2/5} \ell, N^{1/5} L \rightarrow \infty$ in such a way that $A \rightarrow 1/2$, $B \rightarrow I_0$, $D \rightarrow 0$, $\varepsilon^{-3} N \ell^3 \rightarrow 0$, $\varepsilon^{-1} t^{-4} \ell \rightarrow 0$, and $\ell < L$.

Note that from (7) and (21) the limits of A and B will follow if we can prove that $K_1, K_2 \rightarrow 0$. The error term D is of the form

$$D = C t^{-2} (N^{1/5} L)^{-2} + C t^{-4} (N^{2/5} \ell)^{-1} + C (N^{1/5} L)^3 (N^{2/5} \ell)^{-5} + K_3$$

where K_3 is a finite sum of terms of the form

$$C \delta^{-a} \varepsilon^{-b} t^{-c} (N^{2/5} \ell)^d (N^{1/5} L)^e N^{-f}, \quad (70)$$

where $f > 0$. Recall that K_1 and K_2 were sums of expressions of the same form. Note that $\varepsilon^{-3} N \ell^3 = \varepsilon^{-3} (N^{2/5} \ell)^3 N^{-1/5}$, $\varepsilon^{-1} t^{-4} \ell = \varepsilon^{-1} t^{-4} (N^{1/5} \ell) N^{-1/5}$, and $\ell/L = (N^{2/5} \ell) (N^{1/5} L)^{-1} N^{-1/5}$ (which is required to be less than 1) are again all of the form (70).

Now first choose $N^{2/5} \ell \rightarrow \infty$ in such a way that $(N^{2/5} \ell)^d N^{-f} \rightarrow 0$ for all the occurring terms of the form (70). Then choose $N^{1/5} L \rightarrow \infty$ in such a way that $C (N^{1/5} L)^3 (N^{2/5} \ell)^{-5} \rightarrow 0$ and $(N^{2/5} \ell)^d (N^{1/5} L)^e N^{-f} \rightarrow 0$ for all the terms of the form (70). Finally choose $\varepsilon, t, \delta \rightarrow 0$ in such a way that all the terms of the form (70) and the first two terms in D still go to zero. Hence we have achieved all the limits as claimed.

Dyson's formula now follows since by scaling we have that

$$A \int (\nabla \tilde{\Phi})^2 - B \int \tilde{\Phi}^{5/2} = (2A)^{-3/5} (B/I_0)^{4/5} N^{7/5} \left(\frac{1}{2} \int (\nabla \Phi)^2 - I_0 \int \Phi^{5/2} \right),$$

where

$$\Phi(x) = (2A I_0 / B)^{6/5} N^{-8/10} \tilde{\Phi} \left((2A I_0 / B)^{4/5} N^{-1/5} x \right)$$

satisfies $\int \Phi^2 = N^{-1} \int \tilde{\Phi}^2 \leq 1$ if $\int \tilde{\Phi}^2 \leq N$.

Appendix A: Localization of Large Matrices

The following theorem allows us to reduce a big Hermitian matrix, $\mathcal{A}$, to a smaller principal submatrix without changing the lowest eigenvalue very much. The proof can be found in [LS0].

Theorem A.1 (Localization of large matrices). *Suppose that $\mathcal{A}$ is an $N \times N$ Hermitian matrix and let A^k , with $k = 0, 1, \dots, N-1$, denote the matrix consisting of the k^{th} supra- and infra-diagonal of $\mathcal{A}$. Let $\psi \in \mathbb{C}^N$ be a normalized vector and set $d_k = (\psi, A^k \psi)$ and $\lambda = (\psi, \mathcal{A} \psi) = \sum_{k=0}^{N-1} d_k$. (ψ need not be an eigenvector of $\mathcal{A}$.)*

Choose some positive integer $M \leq N$. Then, with M fixed, there is some $n \in [0, N-M]$ and some normalized vector $\phi \in \mathbb{C}^N$ with the property that $\phi_j = 0$ unless $n+1 \leq j \leq n+M$ (i.e., ϕ has length M) and such that

$$(\phi, \mathcal{A} \phi) \leq \lambda + \frac{C}{M^2} \sum_{k=1}^{M-1} k^2 |d_k| + C \sum_{k=M}^{N-1} |d_k|, \quad (71)$$

where $C > 0$ is a universal constant. (Note that the first sum starts with $k = 1$.)

Appendix B: Localization of the Kinetic Energy

Our goal here is to prove a certain lower bound on the many body kinetic energy

$\sum_{i=1}^N -\Delta_i$ acting on the symmetric tensor product space $\bigotimes_S^N L^2(\mathbb{R}^3)$. In order to state the bound we need to introduce some more notation. Let X_0 denote the characteristic function of the cube $[-\ell/2, \ell/2]^3$, in $\mathbb{R}^3$. Let X_z be the characteristic function of the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$, i.e., $X_z(x) = X_0(x - \ell z)$. Let $\mathcal{P}_z$ denote the projection onto the subspace of $L^2(\{z\ell\} + [-\ell/2, \ell/2]^3)$ consisting of functions orthogonal to constants. We shall consider $\mathcal{P}_z$ as a projection in $L^2(\mathbb{R}^3)$. Let $-\Delta_{\text{Neu}}^{(z)}$ be the Neumann Laplacian for the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$. Let $a_0^*(z)$ be the creation operator

$$a_0^*(z) = a^*(\ell^{-3/2} X_z),$$

i.e., the operator creating the constant in the cube $\{z\ell\} + [-\ell/2, \ell/2]^3$. Note that as before $a_0^*(z)$ acts in the Fock space $\bigoplus_{N=0}^{\infty} \bigotimes_S^N L^2(\mathbb{R}^3)$. Products of the form $a_0^*(z) a_0(z')$ or

$a_0(z') a_0^*(z)$ are however bounded operators on the space $\bigotimes_S^N L^2(\mathbb{R}^3)$. Let for all $z \in \mathbb{R}^3$ the function $\chi_z \in C_0^\infty(\{z\ell\} + (-\ell/2, \ell/2)^3)$ be such that $\|\partial^\alpha \chi_z\|_\infty \leq C(\ell t)^{-|\alpha|}$ and $\|\partial^\alpha \sqrt{1 - \chi_z^2}\|_\infty \leq C(\ell t)^{-|\alpha|}$ for some t with $0 < t < 1$ and all multi-indices α with $|\alpha| \leq 3$. We can now state the operator inequality we shall prove. Observe that that we are estimating the one-body kinetic energy operator (the Laplacian) from below by a many-body operator.

Theorem B.1 (Kinetic energy localization). *Let $\Omega \subset \mathbb{R}^3$. Let $w_1, \dots, w_r \in \mathbb{R}^3$ and $\beta_1, \dots, \beta_r > 0$ be such that $X_{w_j} X_0 = 0$ for all $j = 1, \dots, r$ and such that for all $v \in \mathbb{R}^3$ we have that*

$$\sum_{j=1}^r \beta_j (w_j, v)^2 \leq v^2.$$

Ground State Energy of the Two-Component Charged Bose Gas

Then for all $0 < s < t < 1$ and all $\varepsilon > 0$ we have

$$\begin{aligned}
 & \left(1 + \varepsilon + C(s/t)t^{-2} + Cs^{1/2} + Cs \left(\sum_{j=1}^r \beta_j \right)^{1/2} \right) \sum_{i=1}^N -\Delta_i \\
 & \geq \int_{\Omega} \left[\sum_{i=1}^N \mathcal{P}_z^{(i)} \chi_z^{(i)} \frac{(-\Delta_i)^2}{-\Delta_i + (\ell s)^{-2}} \chi_z^{(i)} \mathcal{P}_z^{(i)} + \varepsilon (-\Delta_{i, \text{Neu}}^{(z)}) \right. \\
 & \quad \left. + \sum_{j=1}^r \frac{\beta_j}{\ell^2} \left(\sqrt{a_0^*(z + w_j) a_0(z + w_j) + 1/2} \right. \right. \\
 & \quad \left. \left. - \sqrt{a_0^*(z) a_0(z) + 1/2} \right)^2 \right] dz - \sum_{j=1}^r \frac{\beta_j}{\ell^2} \text{vol}(\Omega). \tag{72}
 \end{aligned}$$

Here all operators are considered in the sense of quadratic forms.

Proof. By rescaling we may consider $\ell = 1$.

Note first that on the quadratic form domain of the Laplacian $-\Delta$ on $\mathbb{R}^3$ we have

$$\int_{\Omega} (-\Delta_{\text{Neu}}^{(z)}) dz \leq \int_{\mathbb{R}^3} (-\Delta_{\text{Neu}}^{(z)}) dz = -\Delta$$

in the sense of quadratic forms.

Let $f_s : \mathbb{R}^3 \rightarrow \mathbb{R}$ be the function $f_s(p) = p^2/(p^2 + s^{-2})$ for some $s > 0$. Then for all $u \in L^2(\mathbb{R}^3)$ we have

$$(2\pi)^{-3} \int f_s(p) |\hat{u}(p)|^2 dp = \int \overline{u(x)} \left(\frac{-\Delta}{-\Delta + s^{-2}} u \right) (x) dx.$$

It follows that for all $u \in L^2(\mathbb{R}^3)$ we have

$$(2\pi)^{-6} \int f_s * |\hat{X}|^2(p) |\hat{u}(p)|^2 dp = \iint \overline{u(x)} X_z(x) \left(\frac{-\Delta}{-\Delta + s^{-2}} (X_z u) \right) (x) dx dz.$$

Lemma B.2. We have

$$\|f_s - (2\pi)^{-3} f_s * |\hat{X}|^2\|_{\infty} \leq Cs^{1/2}.$$

Proof. We calculate

$$|\hat{X}(p)|^2 = \left(\frac{2 \sin(p_1/2)}{p_1} \right)^2 \left(\frac{2 \sin(p_2/2)}{p_2} \right)^2 \left(\frac{2 \sin(p_3/2)}{p_3} \right)^2.$$

In particular $\int_{|p|>r} |\hat{X}(p)|^2 dp \leq Cr^{-1}$ for all $r > 0$. Since $(2\pi)^{-3} \int |\hat{X}|^2(p) dp = 1$, $\|f_s\|_{\infty} \leq 1$, and $\|\nabla f_s\|_{\infty} \leq Cs$ we have

$$\begin{aligned}
 & \left| f_s(p) - (2\pi)^{-3} f_s * |\hat{X}|^2(p) \right| \leq (2\pi)^{-3} \int_{|q|<r} |f_s(p) - f_s(p-q)| |\hat{X}(q)|^2 dq \\
 & \quad + (2\pi)^{-3} \int_{|q|>r} |f_s(p) - f_s(p-q)| |\hat{X}(q)|^2 dq \leq Csr + Cr^{-1}.
 \end{aligned}$$

The result follows from choosing $r = s^{-1/2}$. $\square$

We now write

$$\begin{aligned} p^2 &= (2\pi)^{-3} p^2 f_s * |\hat{X}|^2(p) + p^2 \left(f_s(p) - (2\pi)^{-3} f_s * |\hat{X}|^2(p) \right) + p^2(1 - f_s(p)) \\ &\geq (2\pi)^{-3} p^2 f_s * |\hat{X}|^2(p) - C s^{1/2} p^2 + p^2(1 - f_s(p)). \end{aligned}$$

For all $u \in L^2(\mathbb{R}^3)$ we have

$$\begin{aligned} &(2\pi)^{-6} \int |\hat{u}(p)|^2 p^2 f_s * |\hat{X}|^2(p) dp \\ &= \sum_{i=1}^3 \iint \overline{\partial_i u(x)} X_z(x) \frac{-\Delta}{-\Delta + s^{-2}} (X_z \partial_i u)(x) dx dz. \end{aligned}$$

In other words we have the operator inequality

$$-\Delta \geq \int \mathcal{A}_z dz + \frac{-\Delta s^{-2}}{-\Delta + s^{-2}} - C s^{1/2} (-\Delta), \quad (73)$$

where the operator

$$\mathcal{A}_z = \sum_{i=1}^3 -\partial_i X_z \frac{-\Delta}{-\Delta + s^{-2}} X_z \partial_i$$

is defined as giving the positive quadratic form

$$(u, \mathcal{A}_z u) = \sum_{i=1}^3 \int \overline{\partial_i u(x)} X_z(x) \frac{-\Delta}{-\Delta + s^{-2}} (X_z \partial_i u)(x) dx$$

with domain $H^1(\{z\} + [-1/2, 1/2]^3)$. With the similar notation the operator $\sum_{i=1}^3 -\partial_i X_z \partial_i$ is nothing but the Neumann Laplacian $-\Delta_{\text{Neu}}^{(z)}$.

Lemma B.3. *Given $\theta \in C^\infty(\{z\} + [-1/2, 1/2]^3)$ such that θ is constant near the boundary of the cube $\{z\} + [-1/2, 1/2]^3$. Assume moreover that $\|\partial^\alpha \theta\|_\infty \leq C|t|^{-|\alpha|}$ for some t with $s < t$ and all multi-indices α with $|\alpha| \leq 3$. Then*

$$[[\mathcal{A}_z, \theta], \theta] \geq -C(s/t)(-\Delta_{\text{Neu}}^{(z)}) - C(s/t)t^{-2}X_z.$$

Proof. We calculate

$$\begin{aligned} \left[-\partial_i X_z \frac{-\Delta}{-\Delta + s^{-2}} X_z \partial_i, \theta \right] &= -\partial_i X_z \frac{-\Delta}{-\Delta + s^{-2}} (\partial_i \theta) - \partial_i X_z \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] X_z \partial_i \\ &\quad - (\partial_i \theta) \frac{-\Delta}{-\Delta + s^{-2}} X_z \partial_i \end{aligned}$$

and

$$\begin{aligned} \left[\left[-\partial_i X_z \frac{-\Delta}{-\Delta + s^{-2}} X_z \partial_i, \theta \right], \theta \right] &= -2\partial_i X_z \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] (\partial_i \theta) \\ &\quad - 2(\partial_i \theta) \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] X_z \partial_i - 2(\partial_i \theta) \frac{-\Delta}{-\Delta + s^{-2}} (\partial_i \theta) \\ &\quad - \partial_i X_z \left[\left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right], \theta \right] X_z \partial_i. \end{aligned}$$

Ground State Energy of the Two-Component Charged Bose Gas

We also calculate

$$\left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] = s^{-2} \frac{1}{-\Delta + s^{-2}} [-\Delta, \theta] \frac{1}{-\Delta + s^{-2}}$$

and

$$\begin{aligned} \left[\left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right], \theta \right] &= s^{-2} \frac{1}{-\Delta + s^{-2}} [[-\Delta, \theta], \theta] \frac{1}{-\Delta + s^{-2}} \\ &\quad + 2s^{-2} \frac{1}{-\Delta + s^{-2}} [-\Delta, \theta] \frac{1}{-\Delta + s^{-2}} [\theta, -\Delta] \frac{1}{-\Delta + s^{-2}}. \end{aligned}$$

Hence

$$\left\| \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] \right\| \leq Cs/t \quad \text{and} \quad \left\| \left[\left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right], \theta \right] \right\| \leq C(s/t)^2.$$

Likewise

$$\left\| \left[\left[\frac{-\Delta}{-\Delta + s^{-2}}, \partial_i \theta \right], \partial_i \theta \right] \right\| \leq C(s/t)^2 t^{-2}.$$

Using these estimates we find that

$$\begin{aligned} \partial_i X_z \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] (\partial_i \theta) &+ (\partial_i \theta) \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] X_z \partial_i \\ &\leq -(t/s) \partial_i X_z \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right] \left[\frac{-\Delta}{-\Delta + s^{-2}}, \theta \right]^* X_z \partial_i + (s/t) (\partial_i \theta)^2 \\ &\leq -C(s/t) \partial_i X_z \partial_i + C(s/t) t^{-2} X_z \end{aligned}$$

and

$$\begin{aligned} (\partial_i \theta) \frac{-\Delta}{-\Delta + s^{-2}} (\partial_i \theta) &= \sum_{j=1}^3 -\partial_j (\partial_i \theta) \frac{1}{-\Delta + s^{-2}} (\partial_i \theta) \partial_j \\ &\quad + \sum_{j=1}^3 (\partial_j \partial_i \theta) \frac{\partial_j}{-\Delta + s^{-2}} (\partial_i \theta) - \sum_{j=1}^3 (\partial_i \theta) \frac{\partial_j}{-\Delta + s^{-2}} (\partial_j \partial_i \theta) \\ &\quad - \sum_{j=1}^3 (\partial_j \partial_i \theta) \frac{1}{-\Delta + s^{-2}} (\partial_j \partial_i \theta) \\ &\leq \sum_{j=1}^3 -\partial_j (\partial_i \theta) \frac{1}{-\Delta + s^{-2}} (\partial_i \theta) \partial_j + C(s/t) t^{-2} X_z \\ &\leq -C(s/t)^2 \sum_{j=1}^3 \partial_j X_z \partial_j + C(s/t) t^{-2} X_z. \end{aligned}$$

□

With $\eta_z = \sqrt{1 - \chi_z^2}$ we may write

$$\begin{aligned} \mathcal{A}_z &= \frac{1}{2}(\eta_z^2 + \chi_z^2)\mathcal{A}_z + \frac{1}{2}\mathcal{A}_z(\eta_z^2 + \chi_z^2) \\ &= \chi_z\mathcal{A}_z\chi_z + \eta_z\mathcal{A}_z\eta_z + \frac{1}{2}[[\mathcal{A}_z, \chi_z], \chi_z] + \frac{1}{2}[[\mathcal{A}_z, \eta_z], \eta_z] \\ &\geq \chi_z \frac{(-\Delta)^2}{-\Delta + s^{-2}} \chi_z - C(s/t)(-\Delta_{\text{Neu}}^{(z)}) - C(s/t)t^{-2}X_z. \end{aligned}$$

Here we have used that $\chi_z\mathcal{A}_z\chi_z = \chi_z \frac{(-\Delta)^2}{-\Delta + s^{-2}} \chi_z$, where $-\Delta$ is the Laplacian on $\mathbb{R}^3$.

Clearly we have $\mathcal{P}_z\mathcal{A}_z\mathcal{P}_z = \mathcal{A}_z$. Thus

$$\begin{aligned} \mathcal{A}_z &\geq \mathcal{P}_z\chi_z \frac{(-\Delta)^2}{-\Delta + s^{-2}} \chi_z\mathcal{P}_z - C(s/t)(-\Delta_{\text{Neu}}^{(z)}) - C(s/t)t^{-2}\mathcal{P}_z \\ &\geq \mathcal{P}_z\chi_z \frac{(-\Delta)^2}{-\Delta + s^{-2}} \chi_z\mathcal{P}_z - C(s/t)t^{-2}(-\Delta_{\text{Neu}}^{(z)}), \end{aligned}$$

where we have assumed that $t < 1$ and used that $-\Delta_{\text{Neu}}^{(z)} \geq C\mathcal{P}_z$.

We turn now to the term $-\Delta s^{-2}/(-\Delta + s^{-2})$ in (73).

Lemma B.4. *For all $u \in L^2(\mathbb{R}^3)$ we have that*

$$\sum_{j=1}^r \int_{\mathbb{R}^3} \beta_j |u(x + w_j) - u(x)|^2 dx \leq \left(u, \left(\frac{-\Delta s^{-2}}{-\Delta + s^{-2}} + 2 \left(\sum_{j=1}^r \beta_j \right)^{1/2} s(-\Delta) \right) u \right)$$

Proof. Write $b = 2 \left(\sum_{j=1}^r \beta_j \right)^{1/2}$. On one hand we may write

$$\left(u, \left(\frac{-\Delta s^{-2}}{-\Delta + s^{-2}} + bs(-\Delta) \right) u \right) = (2\pi)^{-3} \int |\hat{u}(p)|^2 \left(\frac{p^2 s^{-2}}{p^2 + s^{-2}} + bsp^2 \right) dp.$$

On the other hand we have

$$\begin{aligned} &\sum_{j=1}^r \int \beta_j |u(x + w_j) - u(x)|^2 dx \\ &= \sum_{j=1}^r (2\pi)^{-3} \int \beta_j |\hat{u}(p)|^2 |\exp(i(p, w_j)/2) - \exp(-i(p, w_j)/2)|^2 dp \\ &= \sum_{j=1}^r (2\pi)^{-3} \int \beta_j |\hat{u}(p)|^2 4 \sin^2((p, w_j)/2) dp. \end{aligned}$$

We thus simply have to prove that

$$\frac{p^2 s^{-2}}{p^2 + s^{-2}} + bsp^2 \geq \sum_{j=1}^r 4\beta_j \sin^2((p, w_j)/2).$$

Ground State Energy of the Two-Component Charged Bose Gas

If $2|p| < b^{1/2}s^{-1/2}$ we have

$$\frac{p^2 s^{-2}}{p^2 + s^{-2}} \geq \frac{p^2}{bs + 1} \geq (1 - bs)p^2 \geq \sum_{j=1}^r 4\beta_j \sin^2((p, w_j)/2) - bsp^2.$$

If $|p| > b^{1/2}s^{-1/2}$ we have

$$bsp^2 \geq b^2 \geq \sum_{j=1}^r 4\beta_j \sin^2((p, w_j)/2).$$

□

Since for all $y \in \mathbb{R}^3$, $(X_{y+\sigma})_{\sigma \in \mathbb{Z}^3}$ is an orthonormal family in $L^2(\mathbb{R}^3)$, we get from Parseval's identity that

$$\int |u(x + w_j) - u(x)|^2 dx \geq \sum_{\sigma \in \mathbb{Z}^3} \left| \int X_{y+\sigma}(x)(u(x + w_j) - u(x)) dx \right|^2.$$

If we integrate this over $y \in [0, 1]^3$ we obtain

$$\int |u(x + w_j) - u(x)|^2 dz \geq \int \left| \int X_z(x)(u(x + w_j) - u(x)) dz \right|^2 dz.$$

The quadratic form on the right of this inequality corresponds to the operator whose second quantization has the form

$$\int (a_0^*(z + w_j) - a_0^*(z))(a_0(z + w_j) - a_0(z)) dz.$$

It is then clear that the main inequality (72) follows from the following lemma.

Lemma B.5. *Let a_1^* and a_2^* be two commuting creation operators. Then*

$$(a_2^* - a_1^*)(a_2 - a_1) \geq \left(\sqrt{a_2^* a_2 + 1/2} - \sqrt{a_1^* a_1 + 1/2} \right)^2 - 1.$$

Proof. We calculate

$$(a_2^* - a_1^*)(a_2 - a_1) = a_2^* a_2 + a_1^* a_1 - (a_2^* a_1 + a_1^* a_2).$$

We have

$$\begin{aligned} (a_2^* a_1 + a_1^* a_2)^2 &\leq (a_2^* a_1 + a_1^* a_2)^2 + (a_2^* a_1 - a_1^* a_2)(a_1^* a_2 - a_2^* a_1) \\ &= 2a_2^* a_1 a_1^* a_2 + 2a_1^* a_2 a_2^* a_1 = 4a_2^* a_2 a_1^* a_1 + 2a_1^* a_1 + 2a_2^* a_2 \\ &= 4(a_2^* a_2 + 1/2)(a_1^* a_1 + 1/2) - 1 \leq 4(a_2^* a_2 + 1/2)(a_1^* a_1 + 1/2). \end{aligned}$$

Since the square root is an operator monotone function we have that

$$a_2^* a_1 + a_1^* a_2 \leq 2\sqrt{a_2^* a_2 + 1/2} \sqrt{a_1^* a_1 + 1/2}.$$

Hence

$$\begin{aligned} (a_2^* - a_1^*)(a_2 - a_1) &\geq a_2^* a_2 + a_1^* a_1 - 2\sqrt{a_2^* a_2 + 1/2} \sqrt{a_1^* a_1 + 1/2} \\ &= \left(\sqrt{a_2^* a_2 + 1/2} - \sqrt{a_1^* a_1 + 1/2} \right)^2 - 1. \end{aligned}$$

□

We shall use Theorem B.1 with $w_1, w_2, \dots, w_r$ being the vectors $\sigma \in \mathbb{Z}^3$ satisfying that $|\sigma|^2 = 2$ or $|\sigma|^2 = 3$. Note that there is a total of 20 such vectors. We define

$$\beta_\sigma = \begin{cases} \frac{1}{12}, & |\sigma|^2 = 2 \\ \frac{1}{24}, & |\sigma|^2 = 3 \end{cases}$$

We then have for all $v \in \mathbb{R}^3$ that

$$\sum_{\substack{\sigma \in \mathbb{Z}^3 \\ |\sigma|=\sqrt{2} \text{ or } |\sigma|=\sqrt{3}}} \beta_\sigma (v, \sigma)^2 = \sum_{\substack{\sigma \in \mathbb{Z}^3 \\ |\sigma|=\sqrt{2}}} \frac{1}{12} (v \cdot \sigma)^2 + \sum_{\substack{\sigma \in \mathbb{Z}^3 \\ |\sigma|=\sqrt{3}}} \frac{1}{24} (v \cdot \sigma)^2 = v^2.$$

References

- [Be] Benguria, R.D.: The Lane-Emden equation revisited. In: *Advances in differential equations and mathematical physics (Birmingham, AL, 2002)*, Contemp. Math. **327**, Providence, RI: Am. Math. Soc., 2003, pp. 11–19
- [B] Bogolubov, N.N.: J. Phys. (U.S.S.R.) **11**, 23 (1947); Bogolubov, N.N., Zubarev, D.N. Sov. Phys. JETP **1**, 83 (1955)
- [CLY] Conlon, J.G., Lieb, E.H., Yau, H-T.: The $N^{7/5}$ law for charged bosons. Commun. Math. Phys. **116**, 417–448 (1988)
- [D] Dyson, F.J.: Ground-state energy of a finite system of charged particles. J. Math. Phys. **8**, 1538–1545 (1967)
- [F] Foldy, L.L.: Charged boson gas, Phys. Rev. **124**, 649–651 (1961); Errata ibid **125**, 2208 (1962)
- [K] Kwong, M.K.: Uniqueness of positive solutions of $\Delta u - u + u^p = 0$ in R^n . Arch. Rat. Mech. Anal. **105**(3), 243–266 (1989)
- [L] Lieb, E.H.: The Bose fluid. In: Brittin, W.E. (eds.), *Lecture Notes in Theoretical Physics VIIC*, Boulder, Co: Univ. of Colorado Press, 1964, pp. 175–224
- [LSo] Lieb, E.H., Solovej, J.P.: Ground State Energy of the One-Component Charged Bose Gas. Commun. Math. Phys. **217** (1), 127–163 (2001) (Erratum: Commun. Math. Phys. **225**, 219–221 (2002))
- [MS] McLeod, K., Serrin, J.: Uniqueness of positive radial solutions of $\Delta u + f(u) = 0$ in R^n . Arch. Rat. Mech. Anal. **99**(2), 115–145 (1987)
- [O] Onsager, L.: Electrostatic Interaction of Molecules. J. Phys. Chem. **43**, 189–196 (1939)
- [S] Solovej, J.P.: Upper Bounds to the Ground State Energies of the One- and Two-Component Charged Bose Gases. arXiv math-ph/0406014
- [Z] Zhang, L.Q.: *Uniqueness of ground state solutions*. Acta Math. Sci. (English Ed.) **8**(4), 449–467 (1988)

Communicated by M. Aizenman

Proof of Bose-Einstein Condensation for Dilute Trapped Gases

Elliott H. Lieb and Robert Seiringer*

Department of Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544
(Received 18 December 2001; published 16 April 2002)

The ground state of bosonic atoms in a trap has been shown experimentally to display Bose-Einstein condensation (BEC). We prove this fact theoretically for bosons with two-body repulsive interaction potentials in the dilute limit, starting from the basic Schrödinger equation; the condensation is 100% into the state that minimizes the Gross-Pitaevskii energy functional. This is the first rigorous proof of BEC in a physically realistic, continuum model.

DOI: 10.1103/PhysRevLett.88.170409

PACS numbers: 05.30.Jp, 03.75.Fi, 67.40.-w

It is gratifying to see the experimental realization, in traps, of the long-predicted Bose-Einstein condensation (BEC) of gases. From the theoretical point of view, however, a rigorous demonstration of this phenomenon—starting from the many-body Hamiltonian of interacting particles—has not yet been achieved. In this Letter, we provide such a rigorous justification for the ground state of 2D or 3D bosons in a trap with repulsive pair potentials, and in the well-defined limit (described below) in which the Gross-Pitaevskii (GP) formula is applicable. It is the first proof of BEC for interacting particles in a continuum (as distinct from lattice) model and in a physically realistic situation.

The difficulty of the problem comes from the fact that BEC is not a consequence of energy considerations alone. The correctness [1] of Bogoliubov's formula for the ground state energy per particle, $e_0(\rho)$, of bosons at low density ρ , namely, $e_0(\rho) = 2\pi\hbar^2\rho a/m$ (with m = particle mass and a = scattering length of the pair potential) shows only that "condensation" exists on local length scales. The same is true [2] in 2D, with Schick's formula [3] $e_0(\rho) = 2\pi\hbar^2\rho/[m|\ln(\rho a^2)|]$. Although it is convenient to assume BEC in the derivation of $e_0(\rho)$, these formulas for $e_0(\rho)$ do not prove BEC. Indeed, in 1D the assumption of BEC leads to a correct formula [4] for $e_0(\rho)$, but there is, presumably, no BEC in 1D ground states [5].

The results just mentioned are for homogeneous gases in the thermodynamic limit. For traps, the GP formula is exact [6,7] in the limit, and one expects BEC into the GP function (instead of into the constant, or zero momentum, function appropriate for the homogeneous gas). This is proven in Theorem 1. In the homogeneous case, the BEC is not 100%, even in the ground state. There is always some depletion. In contrast, BEC in the GP limit is 100% because the $N \rightarrow \infty$ limit is different.

In the homogeneous case, one fixes $a > 0$ and takes $N \rightarrow \infty$ with $\rho = N/\text{volume}$ fixed. For the GP limit, one fixes the external trap potential $V(\mathbf{r})$ and fixes Na , the effective coupling constant, as $N \rightarrow \infty$. A particular, academic example of the trap is $V(\mathbf{r}) = 0$ for $\mathbf{r}$ inside a unit cube and $V(\mathbf{r}) = \infty$ otherwise. By scaling, one can relate this special case to the homogeneous case and thereby

compare the two limits; one sees that the homogeneous case corresponds, mathematically, to the trap case with this special V , but with $Na^3 = \rho a^3$ fixed as $N \rightarrow \infty$. Thus, BEC in the trap case is the easier of the two, reflecting the incompleteness of BEC in the homogeneous case. The lack of depletion in the GP limit is consistent with $\rho a^3 \rightarrow 0$ and with Bogoliubov theory.

We now describe the setting more precisely. We concentrate on the 3D case, and comment on the generalization to 2D at the end of this Letter. The Hamiltonian for N identical bosons in a trap potential V , interacting via a pair potential v , is

$$H = \sum_{i=1}^N [-\Delta_i + V(\mathbf{r}_i)] + \sum_{1 \leq i < j \leq N} v(\mathbf{r}_i - \mathbf{r}_j). \quad (1)$$

It acts on symmetric functions of N variables $\mathbf{r}_i \in \mathbb{R}^3$. Units in which $\hbar^2/2m = 1$ are used. We assume the trap potential V to be a locally bounded function, which tends to infinity as $|\mathbf{r}| \rightarrow \infty$. The interaction potential v is assumed to be non-negative, spherically symmetric, and have a finite scattering length a . (For the definition of scattering length, see [6], [2], or [1].) Note that we do not demand v to be locally integrable; it is allowed to have a hard core, which forces the wave functions to vanish whenever two particles are close together. In the following, we want to let a vary with N , and we do this by scaling; i.e., we write $v(\mathbf{r}) = v_1(\mathbf{r}/a)/a^2$, where v_1 has scattering length 1, and keep v_1 fixed when varying a .

The Gross-Pitaevskii functional is given by

$$\mathcal{E}^{\text{GP}}[\phi] = \int [|\nabla\phi(\mathbf{r})|^2 + V(\mathbf{r})|\phi(\mathbf{r})|^2 + g|\phi(\mathbf{r})|^4] d^3\mathbf{r}.$$

The parameter g is related to the scattering length of the interaction potential appearing in (1) via

$$g = 4\pi Na. \quad (2)$$

We denote by ϕ^{GP} the minimizer of $\mathcal{E}^{\text{GP}}$ under the normalization condition $\int |\phi|^2 = 1$. Existence, uniqueness, and some regularity properties of ϕ^{GP} were proven in the appendix of [6]. In particular, ϕ^{GP} is continuously differentiable and strictly positive. Of course, ϕ^{GP} depends on

g , but we omit this dependence for simplicity of notation. For use later, we define the projector

$$P^{\text{GP}} = |\phi^{\text{GP}}\rangle\langle\phi^{\text{GP}}|. \quad (3)$$

It was shown in [6] (see also Theorem 2 below) that, for each fixed g , the minimization of the GP functional correctly reproduces the large N asymptotics of the ground state energy and density of H —but no assertion about BEC in this limit was made in [6].

BEC in Ψ , the (non-negative and normalized) ground state of H , refers to the reduced one-particle density matrix

$$\gamma(\mathbf{r}, \mathbf{r}') = N \int \Psi(\mathbf{r}, \mathbf{X}) \Psi(\mathbf{r}', \mathbf{X}) d\mathbf{X},$$

where $\mathbf{X} = (\mathbf{r}_2, \dots, \mathbf{r}_N)$ and $d\mathbf{X} = \prod_{j=2}^N d^3\mathbf{r}_j$.

Complete (or 100%) BEC is defined to be the property that $\frac{1}{N}\gamma$ becomes a simple product $f(\mathbf{r})f(\mathbf{r}')$ as $N \rightarrow \infty$, in which case f is called the *condensate wave function*. In the GP limit, i.e., $N \rightarrow \infty$ with $g = 4\pi Na$ fixed, we can show that this is the case, and the condensate wave function is, in fact, the GP minimizer ϕ^{GP} .

THEOREM 1 (Bose-Einstein condensation): *For each fixed g ,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \gamma(\mathbf{r}, \mathbf{r}') = \phi^{\text{GP}}(\mathbf{r}) \phi^{\text{GP}}(\mathbf{r}').$$

Convergence is in the senses that $\text{Trace}[\frac{1}{N}\gamma - P^{\text{GP}}] \rightarrow 0$ and $\int \frac{1}{N} \gamma(\mathbf{r}, \mathbf{r}') - \phi^{\text{GP}}(\mathbf{r}) \phi^{\text{GP}}(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}' \rightarrow 0$.

We remark that Theorem 1 implies that there is 100% condensation for all n -particle reduced density matrices of Ψ ; i.e., they converge to the one-dimensional projector onto the corresponding n -fold product of ϕ^{GP} . To see this, let a^*, a denote the boson creation and annihilation operators for the state ϕ^{GP} , and observe that

$$\begin{aligned} 1 &\geq N^{-n} \langle \Psi | (a^*)^n a^n | \Psi \rangle \approx N^{-n} \langle \Psi | (a^* a)^n | \Psi \rangle \\ &\geq N^{-n} \langle \Psi | a^* a | \Psi \rangle^n \rightarrow 1, \end{aligned}$$

where the terms coming from the commutators $[a, a^*] = 1$ can be neglected since they are of lower order as $N \rightarrow \infty$. The last inequality follows from convexity.

Another corollary, important for the interpretation of experiments, concerns the momentum distribution of the ground state.

COROLLARY 1 (convergence of momentum distribution): *Let $\hat{\rho}(\mathbf{k}) = \int \gamma(\mathbf{r}, \mathbf{r}') \exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')] d^3\mathbf{r} d^3\mathbf{r}'$ denote the one-particle momentum density of Ψ . Then, for each fixed g ,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \hat{\rho}(\mathbf{k}) = |\hat{\phi}^{\text{GP}}(\mathbf{k})|^2,$$

in the sense that $\int |\frac{1}{N} \hat{\rho}(\mathbf{k}) - |\hat{\phi}^{\text{GP}}(\mathbf{k})|^2| d^3\mathbf{k} \rightarrow 0$. Here, $\hat{\phi}^{\text{GP}}$ denotes the Fourier transform of ϕ^{GP} .

Proof.—If $\mathcal{F}$ denotes the (unitary) operator “Fourier transform” and if φ is an arbitrary bounded function with bound $\|\varphi\|_\infty$, then

170409-2

$$\begin{aligned} &\left| \frac{1}{N} \int \hat{\rho} \varphi - \int |\hat{\phi}^{\text{GP}}|^2 \varphi \right| \\ &= |\text{Trace}[\mathcal{F}^*(\gamma/N - P^{\text{GP}}) \mathcal{F} \varphi]| \\ &\leq \|\varphi\|_\infty \text{Trace}|\gamma/N - P^{\text{GP}}|, \end{aligned}$$

whence $\int |\hat{\rho}/N - |\hat{\phi}^{\text{GP}}|^2| \leq \text{Trace}|\gamma/N - P^{\text{GP}}|$. Q.E.D.

Before proving Theorem 1, let us state some prior results on which we shall build. Then we shall outline the proof and formulate two lemmas, which will allow us to prove Theorem 1. We conclude with the proof itself.

Denote by $E^{\text{QM}}(N, a)$ the ground state energy of H and by $E^{\text{GP}}(g)$ the lowest energy of $\mathcal{T}^{\text{GP}}$ with $\int |\phi|^2 = 1$. The following Theorem 2 can be deduced from [6].

THEOREM 2 (asymptotics of energy components): *Let $\rho(\mathbf{r}) = \gamma(\mathbf{r}, \mathbf{r})$ denote the density of the ground state of H . For fixed $g = 4\pi Na$,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} E^{\text{QM}}(N, a) = E^{\text{GP}}(g), \quad (4a)$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{N} \rho(\mathbf{r}) = |\phi^{\text{GP}}(\mathbf{r})|^2, \quad (4b)$$

in the same sense as in Corollary 1. Moreover, if φ_1 denotes the solution to the scattering equation for v_1 (under the boundary condition $\lim_{|\mathbf{r}| \rightarrow \infty} \varphi_1(\mathbf{r}) = 1$) and $s = \int |\nabla \varphi_1|^2 / 4\pi$, then $0 < s \leq 1$ and

$$\begin{aligned} &\lim_{N \rightarrow \infty} \int |\nabla_{\mathbf{r}_1} \Psi(\mathbf{r}_1, \mathbf{X})|^2 d^3\mathbf{r}_1 d\mathbf{X} \\ &= \int |\nabla \phi^{\text{GP}}(\mathbf{r})|^2 d^3\mathbf{r} + g s \int |\phi^{\text{GP}}(\mathbf{r})|^4 d^3\mathbf{r}, \quad (5a) \\ &\lim_{N \rightarrow \infty} \int V(\mathbf{r}_1) |\Psi|^2 d^3\mathbf{r}_1 d\mathbf{X} = \int V(\mathbf{r}) |\phi^{\text{GP}}(\mathbf{r})|^2 d^3\mathbf{r}, \quad (5b) \end{aligned}$$

$$\begin{aligned} &\lim_{N \rightarrow \infty} \frac{1}{2} \sum_{j=2}^N \int v(\mathbf{r}_1 - \mathbf{r}_j) |\Psi(\mathbf{r}_1, \mathbf{X})|^2 d^3\mathbf{r}_1 d\mathbf{X} \\ &= (1 - s) g \int |\phi^{\text{GP}}(\mathbf{r})|^4 d^3\mathbf{r}. \quad (5c) \end{aligned}$$

Only (4) was proven in [6], but (5) follows, as noted in [8], by multiplying V and v by parameters and computing the variation of the energy with respect to them.

[Technical note: The convergence in (4b) was shown in [6] to be in the weak $L^1(\mathbb{R}^3)$ sense, but our result here implies strong convergence, in fact. The proof in Corollary 1, together with Theorem 1 itself, implies this.]

Outline of Proof.—There are two essential ingredients in our proof of Theorem 1. The first is a proof that the part of the kinetic energy that is associated with the interaction v [namely, the second term in (5a)] is mostly located in small balls surrounding each particle. More precisely, these balls can be taken to have radius $N^{-7/17}$, which is

170409-2

much smaller than the mean particle spacing $N^{-1/3}$. This allows us to conclude that the function of $\mathbf{r}$ defined for each fixed value of $\mathbf{X}$ by

$$f_{\mathbf{X}}(\mathbf{r}) = \frac{1}{\phi^{\text{GP}}(\mathbf{r})} \Psi(\mathbf{r}, \mathbf{X}) \geq 0 \quad (6)$$

has the property that $\nabla_{\mathbf{r}} f_{\mathbf{X}}(\mathbf{r})$ is almost zero outside the small balls centered at points of $\mathbf{X}$.

The complement of the small balls has a large volume but it can be a weird set; it need not even be connected. Therefore, the smallness of $\nabla_{\mathbf{r}} f_{\mathbf{X}}(\mathbf{r})$ in this set does not guarantee that $f_{\mathbf{X}}(\mathbf{r})$ is nearly constant (in $\mathbf{r}$), or even that it is continuous. We need $f_{\mathbf{X}}(\mathbf{r})$ to be nearly constant in order to conclude BEC. What saves the day is the knowledge that the total kinetic energy of $f_{\mathbf{X}}(\mathbf{r})$ (including the balls) is not huge. The result that allows us to combine these two pieces of information in order to deduce the almost constancy of $f_{\mathbf{X}}(\mathbf{r})$ is the generalized Poincaré inequality in Lemma 2. (End of outline.)

Using the results of Theorem 2, partial integration, and the GP equation [i.e., the variational equation for ϕ^{GP} , see [6], Eq. 2.4)], we see that

$$\lim_{N \rightarrow \infty} \int |\phi^{\text{GP}}(\mathbf{r})|^2 |\nabla_{\mathbf{r}} f_{\mathbf{X}}|^2 d^3 \mathbf{r} d\mathbf{X} = g_s \int |\phi^{\text{GP}}|^4 d^3 \mathbf{r}. \quad (7)$$

The following lemma shows that to leading order all the energy in (7) is concentrated in small balls.

LEMMA 1 (localization of energy): For fixed $\mathbf{X}$ let

$$\Omega_{\mathbf{X}} = \{\mathbf{r} \in \mathbf{R}^3 | \min_{k \geq 2} |\mathbf{r} - \mathbf{r}_k| \geq N^{-7/17}\}. \quad (8)$$

Then

$$\lim_{N \rightarrow \infty} \int d\mathbf{X} \int_{\Omega_{\mathbf{X}}} d^3 \mathbf{r} |\phi^{\text{GP}}(\mathbf{r})|^2 |\nabla_{\mathbf{r}} f_{\mathbf{X}}(\mathbf{r})|^2 = 0.$$

Proof.—We shall show that, as $N \rightarrow \infty$,

$$\begin{aligned} & \int d\mathbf{X} \int_{\Omega_{\mathbf{X}}} d^3 \mathbf{r} |\phi^{\text{GP}}(\mathbf{r})|^2 |\nabla_{\mathbf{r}} f_{\mathbf{X}}(\mathbf{r})|^2 \\ & + \frac{1}{2} \left[\int d\mathbf{X} \int d^3 \mathbf{r} |\phi^{\text{GP}}(\mathbf{r})|^2 \right. \\ & \quad \times \sum_{k \geq 2} v(\mathbf{r} - \mathbf{r}_k) |f_{\mathbf{X}}(\mathbf{r})|^2 \left. \right] \\ & - 2g \int d\mathbf{X} \int d^3 \mathbf{r} |\phi^{\text{GP}}(\mathbf{r})|^4 |f_{\mathbf{X}}(\mathbf{r})|^2 \\ & \geq -g \int |\phi^{\text{GP}}(\mathbf{r})|^4 d^3 \mathbf{r} - o(1), \quad (9) \end{aligned}$$

which implies the assertion of the lemma by virtue of (7) and the results of Theorem 2. Here, $\Omega_{\mathbf{X}}$ is the complement of $\Omega_{\mathbf{X}}$. The proof of (9) is actually just a detailed examination of the lower bounds to the energy derived in [6] and

[1], and we use the methods in [1,6], just describing the differences from the case considered here.

Writing $f_{\mathbf{X}}(\mathbf{r}) = \prod_{k \geq 2} \phi^{\text{GP}}(\mathbf{r}_k) F(\mathbf{r}, \mathbf{X})$ and using the fact that F is symmetric in the particle coordinates, we see that (9) is equivalent to

$$\frac{1}{N} Q(F) \geq -g \int |\phi^{\text{GP}}|^4 - o(1), \quad (10)$$

where Q is the quadratic form

$$\begin{aligned} Q(F) = & \sum_{i=1}^N \int_{\Omega_i^c} |\nabla_i F|^2 \prod_{k=1}^N |\phi^{\text{GP}}(\mathbf{r}_k)|^2 d^3 \mathbf{r}_k \\ & + \sum_{1 \leq i < j \leq N} v(\mathbf{r}_i - \mathbf{r}_j) |F|^2 \prod_{k=1}^N |\phi^{\text{GP}}(\mathbf{r}_k)|^2 d^3 \mathbf{r}_k \\ & - 2g \sum_{i=1}^N \int |\phi^{\text{GP}}(\mathbf{r}_i)|^2 |F|^2 \prod_{k=1}^N |\phi^{\text{GP}}(\mathbf{r}_k)|^2 d^3 \mathbf{r}_k, \quad (11) \end{aligned}$$

with $\Omega_i^c = \{(\mathbf{r}_1, \mathbf{X}) \in \mathbf{R}^{3N} | \min_{k \neq i} |\mathbf{r}_i - \mathbf{r}_k| \leq N^{-7/17}\}$.

While (10) is not true for all conceivable F 's satisfying the condition $\int |F|^2 \prod_{k=1}^N |\phi^{\text{GP}}(\mathbf{r}_k)|^2 d^3 \mathbf{r}_k = 1$, it is true for an F , such as ours, that has bounded kinetic energy (7). Equations (4.11), (4.12), and (4.23)–(4.25), proved in [6], are similar to (10) and (11) and almost establish (10), but there are two differences which we now explain.

(i) In our case, the kinetic energy of particle i is restricted to the subset of $\mathbf{R}^{3N}$ in which $\min_{k \neq i} |\mathbf{r}_i - \mathbf{r}_k| \leq N^{-7/17}$. However, from the proof of the lower bound to the ground state energy of a homogeneous Bose gas derived in [1] [especially Lemma 1 and Eq. (26) there], which enters the calculations in [6], we see that only this part of the kinetic energy enters the proof of the lower bound—except for some additional piece with a relative magnitude $\varepsilon = O(N^{-2/17})$. In the notation of [1], the radius of the balls used in the application of Lemma 1 is chosen to be $R = aY^{-5/17}$, which, in the GP regime, is $R = O(N^{-7/17})$ since, for fixed Na , $Y = O(a^3 N) = O(N^{-2})$. (See [9] for a more complete discussion about the choice of R .) The *a priori* knowledge that the total kinetic energy is bounded by (7) tells us that the additional piece, which is ε times the total kinetic energy, is truly $O(\varepsilon)$ and goes to zero as $N \rightarrow \infty$.

(ii) In [6] all integrals were restricted to some arbitrarily big, but finite, box of size R' . However, the difference in the energy is easily estimated to be smaller than $2gN \max_{|\mathbf{r}| \geq R'} |\phi^{\text{GP}}(\mathbf{r})|^2$, which, divided by N , is arbitrarily small, since $\phi^{\text{GP}}(\mathbf{r})$ decreases faster than exponentially at infinity ([6], Lemma A.5).

Proceeding exactly as in [6] and taking the differences (i) and (ii) into account, we arrive at (10). Q.E.D.

In the following, $\mathcal{K} \subset \mathbf{R}^m$ denotes a bounded and connected set that is sufficiently nice so that the Poincaré-Sobolev inequality (see [10], Theorem 8.12) holds on $\mathcal{K}$. In particular, this is the case if $\mathcal{K}$ satisfies the cone property [10] (e.g., if $\mathcal{K}$ is a ball or a cube).

We introduce the general notation that $f \in L^p(\mathcal{K})$ if the norm $\|f\|_{L^p(\mathcal{K})} = [\int_{\mathcal{K}} |f(\mathbf{r})|^p d^m \mathbf{r}]^{1/p}$ is finite.

LEMMA 2 (generalized Poincaré inequality): For $m \geq 2$ let $\mathcal{K} \subset \mathbb{R}^m$ be as explained above, and let h be a bounded function with $\int_{\mathcal{K}} h = 1$. There exists a constant C (depending only on $\mathcal{K}$ and h) such that, for all sets $\Omega \subset \mathcal{K}$ and all $f \in H^1(\mathcal{K})$ [i.e., $f \in L^2(\mathcal{K})$ and $\nabla f \in L^2(\mathcal{K})$] with $\int_{\mathcal{K}} f h d^m \mathbf{r} = 0$, the inequality

$$\begin{aligned} \int_{\Omega} |\nabla f(\mathbf{r})|^2 d^m \mathbf{r} + \left(\frac{|\Omega^c|}{|\mathcal{K}|} \right)^{2/m} \int_{\mathcal{K}} |\nabla f(\mathbf{r})|^2 d^m \mathbf{r} \\ \geq \frac{1}{C} \int_{\mathcal{K}} |f(\mathbf{r})|^2 d^m \mathbf{r} \quad (12) \end{aligned}$$

holds. Here $|\cdot|$ is the volume of a set, and $\Omega^c = \mathcal{K} \setminus \Omega$.

Proof.—By the usual Poincaré-Sobolev inequality on $\mathcal{K}$ (see [10], Theorem 8.12),

$$\begin{aligned} \|f\|_{L^2(\mathcal{K})}^2 &\leq \tilde{C} \|\nabla f\|_{L^{2m/(m+2)}(\mathcal{K})}^2 \\ &\leq 2\tilde{C} (\|\nabla f\|_{L^{2m/(m+2)}(\Omega)}^2 + \|\nabla f\|_{L^{2m/(m+2)}(\Omega^c)}^2), \end{aligned}$$

if $m \geq 2$ and $\int_{\mathcal{K}} f h = 0$. Applying Hölder's inequality

$$\|\nabla f\|_{L^{2m/(m+2)}(\Omega)} \leq \|\nabla f\|_{L^2(\Omega)} |\Omega|^{1/m}$$

(and the analogue with Ω replaced by Ω^c), we see that (12) holds with $C = 2|\mathcal{K}|^{2/m} \tilde{C}$. Q.E.D.

The important point in Lemma 2 is that there is no restriction on Ω concerning regularity or connectivity.

Proof of Theorem 1.—For some $R > 0$ let $\mathcal{K} = \{\mathbf{r} \in \mathbb{R}^3, |\mathbf{r}| \leq R\}$, and define

$$\langle f_{\mathbf{x}} \rangle_{\mathcal{K}} = \frac{1}{\int_{\mathcal{K}} |\phi^{\text{GP}}(\mathbf{r})|^2 d^3 \mathbf{r}} \int_{\mathcal{K}} |\phi^{\text{GP}}(\mathbf{r})|^2 f_{\mathbf{x}}(\mathbf{r}) d^3 \mathbf{r}.$$

We shall use Lemma 2, with $m = 3$, $h(\mathbf{r}) = |\phi^{\text{GP}}(\mathbf{r})|^2 / \int_{\mathcal{K}} |\phi^{\text{GP}}|^2$, $\Omega = \Omega_{\mathbf{x}} \cap \mathcal{K}$, and $f(\mathbf{r}) = f_{\mathbf{x}}(\mathbf{r}) - \langle f_{\mathbf{x}} \rangle_{\mathcal{K}}$ [see (8) and (6)]. Since ϕ^{GP} is bounded on $\mathcal{K}$ above and below by some positive constants, this lemma also holds (with a different constant C') with $d^3 \mathbf{r}$ replaced by $|\phi^{\text{GP}}(\mathbf{r})|^2 d^3 \mathbf{r}$ in (12). Therefore,

$$\begin{aligned} \int d\mathbf{X} \int_{\mathcal{K}} d^3 \mathbf{r} |\phi^{\text{GP}}(\mathbf{r})|^2 [f_{\mathbf{x}}(\mathbf{r}) - \langle f_{\mathbf{x}} \rangle_{\mathcal{K}}]^2 \\ \leq C' \int d\mathbf{X} \left[\int_{\Omega_{\mathbf{x}} \cap \mathcal{K}} |\phi^{\text{GP}}(\mathbf{r})|^2 |\nabla_{\mathbf{r}} f_{\mathbf{x}}(\mathbf{r})|^2 d^3 \mathbf{r} \right. \\ \left. + \frac{N^{-8/51}}{R^2} \int_{\mathcal{K}} |\phi^{\text{GP}}(\mathbf{r})|^2 |\nabla_{\mathbf{r}} f_{\mathbf{x}}(\mathbf{r})|^2 d^3 \mathbf{r} \right], \quad (13) \end{aligned}$$

where we used $|\Omega_{\mathbf{x}}^c \cap \mathcal{K}| \leq (4\pi/3)N^{-4/17}$. The first integral on the right side of (13) tends to zero as $N \rightarrow \infty$ by Lemma 1, and the second is bounded by (7). We conclude, since $\int_{\mathcal{K}} |\phi^{\text{GP}}(\mathbf{r})|^2 f_{\mathbf{x}}(\mathbf{r}) d^3 \mathbf{r} \leq \int_{\mathbb{R}^3} |\phi^{\text{GP}}(\mathbf{r})|^2 f_{\mathbf{x}}(\mathbf{r}) d^3 \mathbf{r}$, that

170409-4

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N} \langle \phi^{\text{GP}} | \gamma | \phi^{\text{GP}} \rangle \\ \geq \int_{\mathcal{K}} |\phi^{\text{GP}}(\mathbf{r})|^2 d^3 \mathbf{r} \\ \times \lim_{N \rightarrow \infty} \int d\mathbf{X} \int_{\mathcal{K}} d^3 \mathbf{r} |\Psi(\mathbf{r}, \mathbf{X})|^2. \end{aligned}$$

It follows from (4b) that the right side of this inequality equals $[\int_{\mathcal{K}} |\phi^{\text{GP}}(\mathbf{r})|^2 d^3 \mathbf{r}]^2$. Since the radius of $\mathcal{K}$ was arbitrary, $\frac{1}{N} \langle \phi^{\text{GP}} | \gamma | \phi^{\text{GP}} \rangle \rightarrow 1$, implying Theorem 1 (cf. [11], Theorem 2.20). Q.E.D.

We remark that the method presented here also works in the case of a 2D Bose gas. The relevant parameter to be kept fixed in the GP limit is $g = 4\pi N / |\ln(a^2 N)|$; all other considerations carry over without essential change, using the results in [2,7]. A minor difference concerns the parameter s in Theorem 2, which can be shown to be always equal to 1 in 2D; i.e., the interaction energy is purely kinetic in the GP limit (see [12]). We also point out that our method necessarily fails for the 1D Bose gas, where there is presumably no BEC [5]. An analogue of Lemma 1 cannot hold in the 1D case since even a hard core potential with an arbitrarily small range produces an interaction energy that is not localized on scales smaller than the mean particle spacing. There is also no GP limit for the one-dimensional Bose gas in the above sense.

We are grateful to Jakob Yngvason for helpful discussions. E. H. L. was partially supported by the U.S. National Science Foundation Grant No. PHY 98 20650. R. S. was supported by the Austrian Science Foundation.

*On leave from Institut für Theoretische Physik, Universität Wien, Boltzmannsgasse 5, 1090 Vienna, Austria.

- [1] E. H. Lieb and J. Yngvason, Phys. Rev. Lett. **80**, 2504 (1998).
- [2] E. H. Lieb and J. Yngvason, J. Stat. Phys. **103**, 509 (2001).
- [3] M. Schick, Phys. Rev. A **3**, 1067 (1971).
- [4] E. H. Lieb and W. Liniger, Phys. Rev. **130**, 1605 (1963).
- [5] L. Pitaevskii and S. Stringari, J. Low Temp. Phys. **85**, 377 (1991).
- [6] E. H. Lieb, R. Seiringer, and J. Yngvason, Phys. Rev. A **61**, 043602 (2000).
- [7] E. H. Lieb, R. Seiringer, and J. Yngvason, Commun. Math. Phys. **224**, 17 (2001).
- [8] A. Y. Cherny and A. A. Shalenko, Phys. Lett. A **293**, 287 (2002).
- [9] E. H. Lieb and J. Yngvason, in *Differential Equations and Mathematical Physics: Proceedings of the 1999 Conference at the University of Alabama*, edited by R. Weikard and G. Weinstein (International Press, Providence, 2000), p. 295.
- [10] E. H. Lieb and M. Loss, *Analysis* (American Mathematical Society, Providence, 2001), 2nd ed.
- [11] B. Simon, *Trace Ideals and Their Application* (Cambridge University, Cambridge, England, 1979).
- [12] A. Y. Cherny and A. A. Shalenko, Phys. Rev. E **64**, 027105 (2001).

170409-4

PHYSICAL REVIEW B 66, 134529 (2002)

Superfluidity in dilute trapped Bose gases

Elliott H. Lieb* and Robert Seiringer†

Department of Physics, Jadwin Hall, Princeton University, P. O. Box 708, Princeton, New Jersey 08544

Jakob Yngvason‡

Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, A-1090 Vienna, Austria

(Received 28 May 2002; revised manuscript received 5 August 2002; published 31 October 2002)

A commonly used theoretical definition of superfluidity in the ground state of a Bose gas is based on the response of the system to an imposed velocity field or, equivalently, to twisted boundary conditions in a box. We are able to carry out this program in the case of a dilute interacting Bose gas in a trap, and we prove that a gas with repulsive interactions is 100% superfluid in the dilute limit in which the Gross-Pitaevskii equation is exact. This is the first example in an experimentally realistic continuum model in which superfluidity is rigorously verified.

DOI: 10.1103/PhysRevB.66.134529

PACS number(s): 05.30.Jp, 03.75.Fi, 67.40.-w

I. INTRODUCTION

The phenomenological two-fluid model of superfluidity (see, e.g., Ref. 1) is based on the idea that the particle density ρ is composed of two parts, the density ρ_s of the inviscid superfluid and the normal fluid density ρ_n . If an external velocity field is imposed on the fluid (for instance by moving the walls of the container) only the viscous normal component responds to the velocity field, while the superfluid component stays at rest. In accord with these ideas the superfluid density in the ground state is often defined as follows²: Let E_0 denote the ground state energy of the system in the rest frame and E'_0 the ground state energy, measured in the moving frame, when a velocity field $\mathbf{v}$ is imposed. Then, for small $\mathbf{v}$,

$$\frac{E'_0}{N} = \frac{E_0}{N} + (\rho_s/\rho) \frac{1}{2} m \mathbf{v}^2 + O(|\mathbf{v}|^4), \quad (1)$$

where N is the particle number and m the particle mass. At positive temperatures the ground state energy should be replaced by the free energy. [Remark: It is important here that Eq. (1) holds uniformly for all large N , i.e., that the error term $O(|\mathbf{v}|^4)$ can be bounded independently of N . For fixed N and a finite box, Eq. (1) with $\rho_s/\rho=1$ always holds for a Bose gas with an arbitrary interaction if $\mathbf{v}$ is small enough, owing to the discreteness of the energy spectrum.³] There are other definitions of the superfluid density that may lead to slightly different results,⁴ but this is the one we shall use in this paper. We shall not dwell on this issue here, since it is not clear that there is a “one-size-fits-all” definition of superfluidity. For instance, in the definition we use here the ideal Bose gas is a perfect superfluid in its ground state, whereas the definition of Landau in terms of a linear dispersion relation of elementary excitations would indicate otherwise. We emphasize that we are not advocating any particular approach to the superfluidity question; our contribution here consists of taking one standard definition and making its consequences explicit.

One of the unresolved issues in the theory of superfluidity is its relation to Bose-Einstein condensation (BEC). It has been argued that in general neither condition is necessary for

the other (cf., e.g., Refs. 5–7). A simple example illustrating the fact that BEC is not necessary for superfluidity is the one-dimensional hard-core Bose gas. This system is well known to have a spectrum like that of an ideal Fermi gas,⁸ and it is easy to see that it is superfluid in its ground state in the sense of Eq. (1). On the other hand, it has no BEC.^{9,10} The definition of the superfluid velocity as the gradient of the phase of the condensate wave function^{2,11} is clearly not applicable in such cases.

We do not give a historical overview of superfluidity because excellent review articles are available.^{11,12} While the early investigations of superfluidity and Bose-Einstein condensation were mostly concerned with liquid Helium 4, it has become possible in recent years to study these phenomena in dilute trapped gases of alkali atoms.¹³ The experimental success in realizing BEC in such gases has led to a large number of theoretical papers on this subject. Most of these works take BEC for granted, and start off with the Gross-Pitaevskii (GP) equation to describe the condensate wave function. A rigorous justification of these assumptions is however a difficult task, and only very recently BEC has been rigorously proved for a physically realistic many-body Hamiltonian.¹⁴ It is clearly of interest to show that superfluidity also holds in this model, and this is what we accomplish here. We prove that the ground state of a Bose gas with short range, repulsive interaction is 100% superfluid in the dilute limit in which the Gross-Pitaevskii description of the gas is exact. This is the limit in which the particle number tends to infinity, but the ratio Na/L , where a is the scattering length of the interaction potential and L the box size, is kept fixed. (The significance of the parameter Na/L is that it is the ratio of the ground state energy per particle, $\sim Na/L^3$, to the lowest excitation energy in the box, $\sim 1/L^2$.) In addition we show that the gas remains 100% Bose-Einstein condensed in this limit, also for a finite velocity $\mathbf{v}$. Both results can be generalized from periodic boxes to (nonconstant) velocity fields in a cylindrical geometry.

The results of this paper have been conjectured for many years, and it is gratifying that they can be proved from first principles. To our knowledge they represent the first example of a rigorous verification of superfluidity in an experimentally realistic continuum model.

ELLIOTT H. LIEB, ROBERT SEIRINGER, AND JAKOB YNGVASON

PHYSICAL REVIEW B 66, 134529 (2002)

We wish to emphasize that in this GP limit the fact that there is 100% condensation does not mean that no significant interactions occur. The kinetic and potential energies can differ markedly from that obtained with a simple variational function that is an N -fold product of one-body condensate wave functions. This assertion might seem paradoxical, and the explanation is that near the GP limit the region in which the wave function differs from the condensate function has a tiny volume that goes to zero as $N \rightarrow \infty$. Nevertheless, the interaction energy, which is proportional to N , resides in this tiny volume.

II. SETTING AND MAIN RESULTS

We consider a Bose gas with the Hamiltonian

$$H_N = -\mu \sum_{j=1}^N \nabla_j^2 + \sum_{1 \leq i < j \leq N} v(|\mathbf{r}_i - \mathbf{r}_j|), \quad (2)$$

where $\mu = \hbar^2/(2m)$ and the interaction potential v is non-negative and of finite range. The two-body scattering length of v is denoted by a . The Hamiltonian acts on totally symmetric functions Ψ of N variables $\mathbf{r}_i = (x_i, y_i, z_i) \in \mathcal{K} \subset \mathbb{R}^3$, where $\mathcal{K}$ denotes the cube $[0, L]^3$ of side length L . (We could easily use a cuboid of sides L_1, L_2, L_3 instead.) We assume periodic boundary conditions in all three coordinate directions.

Imposing an external velocity field $\mathbf{v} = (0, 0, \pm|\mathbf{v}|)$ means that the momentum operator $\mathbf{p} = -i\hbar\nabla$ is replaced by $\mathbf{p} - m\mathbf{v}$, retaining the periodic boundary conditions. The Hamiltonian in the moving frame is thus

$$H'_N = -\mu \sum_{j=1}^N (\nabla_j + i\varphi/L)^2 + \sum_{1 \leq i < j \leq N} v(|\mathbf{r}_i - \mathbf{r}_j|), \quad (3)$$

where $\varphi = (0, 0, \varphi)$ and the dimensionless phase φ is connected to the velocity $\mathbf{v}$ by

$$\varphi = \frac{\pm|\mathbf{v}|Lm}{\hbar}. \quad (4)$$

Let $E_0(N, a, \varphi)$ denote the ground state energy of Eq. (3) with periodic boundary conditions. Obviously it is no restriction to consider only the case $-\pi \leq \varphi \leq \pi$, since E_0 is periodic in φ with period 2π . For Ψ_0 the ground state of H'_N , let γ_N be its one-particle reduced density matrix

$$\begin{aligned} \gamma_N(\mathbf{r}, \mathbf{r}') &= N \int_{\mathcal{K}^{N-1}} \Psi_0(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \\ &\quad \times \Psi_0^*(\mathbf{r}', \mathbf{r}_2, \dots, \mathbf{r}_N) d\mathbf{r}_2 \cdots d\mathbf{r}_N. \end{aligned} \quad (5)$$

We are interested in the GP limit $N \rightarrow \infty$ with Na/L fixed. We also fix the box size L . This means that a should vary like $1/N$ which can be achieved by writing $v(\mathbf{r}) = a^{-2}v_1(\mathbf{r}/a)$, where v_1 is a fixed potential with scattering length 1, while a changes with N .

THEOREM 1 (*Superfluidity of homogeneous gas*). For $|\varphi| \leq \pi$

$$\lim_{N \rightarrow \infty} \frac{E_0(N, a, \varphi)}{N} = 4\pi\mu a \rho + \mu \frac{\varphi^2}{L^2} \quad (6)$$

in the limit $N \rightarrow \infty$ with Na/L and L fixed. Here $\rho = N/L^3$, so $a\rho$ is fixed too. In the same limit, for $|\varphi| < \pi$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \gamma_N(\mathbf{r}, \mathbf{r}') = \frac{1}{L^3} \quad (7)$$

in trace class norm, i.e.,

$$\lim_{N \rightarrow \infty} \text{Tr}[\gamma_N/N - |L^{-3/2}\rangle\langle L^{-3/2}|] = 0.$$

Note that, by definition (1) of ρ_s and Eq. (4), Eq. (6) means that $\rho_s = \rho$, i.e., there is 100% superfluidity. For $\varphi = 0$, Eq. (6) was first proved in Ref. 15. Eq. (7) for $\varphi = 0$ is the BEC proved in Ref. 14.

Remarks. (1) By a unitary gauge transformation,

$$(U\Psi)(\mathbf{r}_1, \dots, \mathbf{r}_N) = e^{i\varphi(\sum_i z_i)/L} \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N), \quad (8)$$

the passage from Eq. (2) to Eq. (3) is equivalent to replacing periodic boundary conditions in a box by the *twisted boundary condition*

$$\Psi(\mathbf{r}_1 + (0, 0, L), \mathbf{r}_2, \dots, \mathbf{r}_N) = e^{i\varphi} \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (9)$$

in the direction of the velocity field, while retaining the original Hamiltonian [Eq. (2)].

(2) The criterion $|\varphi| \leq \pi$ means that $|\mathbf{v}| \leq \pi\hbar/(mL)$. The corresponding energy $\frac{1}{2}m[\pi\hbar/(mL)]^2$ is the gap in the excitation spectrum of the one-particle Hamiltonian in the finite-size system.

(3) The reason that we have to restrict ourselves to $|\varphi| < \pi$ in the second part of theorem 1 is that for $|\varphi| = \pi$ there are two ground states of the operator $(\nabla + i\varphi/L)^2$ with periodic boundary conditions. All we can say in this case is that there is a subsequence of γ_N that converges to a density matrix of rank ≤ 2 , whose range is spanned by these two functions.

Theorem 1 can be generalized in various ways to a physically more realistic setting. As an example, let C be a finite cylinder based on an annulus centered at the origin. Given a bounded, real function $a(r, z)$ let A be the vector field (in polar coordinates) $A(r, \theta, z) = \varphi a(r, z) \hat{e}_\theta$, where $\hat{e}_\theta$ is the unit vector in the θ direction. We also allow for a bounded external potential $V(r, z)$ that does not depend on θ .

Using the methods of Appendix A in Ref. 16, it is not difficult to see that there exists a $\varphi_0 > 0$, depending only on C and $a(r, z)$, such that for all $|\varphi| < \varphi_0$ there is a unique minimizer ϕ^{GP} of the Gross-Pitaevskii functional

$$\begin{aligned} \mathcal{E}^{\text{GP}}[\phi] &= \int_C \{ \mu [|\nabla + iA(\mathbf{r})|\phi(\mathbf{r})|^2 + V(\mathbf{r})|\phi(\mathbf{r})|^2 \\ &\quad + 4\pi\mu Na|\phi(\mathbf{r})|^4] d^3\mathbf{r} \end{aligned} \quad (10)$$

under the normalization condition $\int |\phi|^2 = 1$. This minimizer does not depend on θ , and can be chosen to be positive, for the following reason: The relevant term in the kinetic energy is $T = -r^{-2}[\partial/\partial\theta + i\varphi a(r, z)]^2$. If $|\varphi a(r, z)| < 1/2$, it is

easy to see that $T \geq \varphi^2 a(r, z)^2$, in which case, without raising the energy, we can replace ϕ by the square root of the θ -average of $|\phi|^2$. This can only lower the kinetic energy¹⁷ and, by convexity of $x \rightarrow x^2$, this also lowers the ϕ^4 term.

We denote the ground state energy of $\mathcal{E}^{\text{GP}}$ by E^{GP} , depending on Na and φ . The following theorem 2 concerns the ground state energy E_0 of

$$H_N^A = \sum_{j=1}^N \{-\mu[\nabla_j + iA(\mathbf{r}_j)]^2 + V(\mathbf{r}_j)\} + \sum_{1 \leq i < j \leq N} v(|\mathbf{r}_i - \mathbf{r}_j|), \quad (11)$$

with Neumann boundary conditions on $\mathcal{C}$, and the one-particle reduced density matrix γ_N of the ground state, respectively. Different boundary conditions can be treated in the same manner, if they are also used in Eq. (10).

Remark. As a special case, consider a uniformly rotating system. In this case $A(\mathbf{r}) = \varphi r \hat{e}_\theta$, where 2φ is the angular velocity. H_N^A is the Hamiltonian in the rotating frame, but with external potential $V(\mathbf{r}) + \mu A(\mathbf{r})^2$ [see, e.g., Ref. 11 (p. 131)].

THEOREM 2 (*Superfluidity in a cylinder*). For $|\varphi| < \varphi_0$

$$\lim_{N \rightarrow \infty} \frac{E_0(N, a, \varphi)}{N} = E^{\text{GP}}(Na, \varphi) \quad (12)$$

in the limit $N \rightarrow \infty$ with Na fixed. In the same limit,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \gamma_N(\mathbf{r}, \mathbf{r}') = \phi^{\text{GP}}(\mathbf{r}) \phi^{\text{GP}}(\mathbf{r}') \quad (13)$$

in trace class norm, i.e., $\lim_{N \rightarrow \infty} \text{Tr}[|\gamma_N/N - |\phi^{\text{GP}}\rangle\langle\phi^{\text{GP}}||] = 0$.

In the case of a uniformly rotating system, where 2φ is the angular velocity, the condition $|\varphi| < \varphi_0$ in particular means that the angular velocity is smaller than the critical velocity for creating vortices.¹⁸

Remark. In the special case of the curl-free vector potential $A(r, \theta) = \varphi r^{-1} \hat{e}_\theta$, i.e., $a(r, z) = r^{-1}$, one can say more about the role of φ_0 . In this case, there is a unique GP minimizer for all $\varphi \in \mathbb{Z} + \frac{1}{2}$, whereas there are two minimizers for $\varphi \in \mathbb{Z} + \frac{1}{2}$. Part two of theorem 2 holds in this special case for all $\varphi \in \mathbb{Z} + \frac{1}{2}$, and (12) is true even for all φ .

III. PROOFS

In the following, we will present only a proof of theorem 1 for simplicity. Theorem 2 can be proved using the same methods, and additionally the methods of Ref. 14 to deal with the inhomogeneity of the system.

Before giving the formal proofs, we outline the main ideas. The strategy is related to the one in Ref. 14, but requires substantial generalizations of the techniques. A crucial element of the proof, stated in lemma 1 below, is the fact that the interaction energy can be localized in small balls around each particle. This part uses a lemma of Dyson,¹⁹ and its generalization in Ref. 15, which converts a strong short range potential into a soft potential. This lemma can be also be applied to the case of an external velocity field, i.e., a

$U(1)$ gauge field in the kinetic term of the Hamiltonian, owing to the “diamagnetic inequality.”¹⁷ This inequality says that the additional gauge field increases the kinetic energy density.

The second main part of the proof is the generalized Poincaré inequality given in lemma 2. We recall that an essential ingredient of the proof of Bose-Einstein condensation in Ref. 14 was showing that the fact that the kinetic energy density is small in most of the configuration space implies that the one-body reduced density matrix is essentially constant. The difficulty comes from the fact that the region in which the kinetic energy is small can, in principle, be broken up into disjoint subregions, thereby permitting different constants in different subregions. The fact that this does not happen is the content of the generalized Poincaré inequality. In the present case we have an additional complication coming from the imposed gauge field. The old Poincaré inequality does not suffice; one now has to measure the kinetic energy density relative to the lowest energy of a free particle in the gauge field rather than to zero. This is an essential complication. While the previous (generalized) Poincaré inequality could, after some argumentation, be related to the standard Poincaré inequality,¹⁷ this one, with the gauge field, requires a different proof.

Proof of Theorem 1. As in Ref. 15 we define $Y = (4\pi/3)\rho a^3$. Note that in the limit considered, $Y \sim N^{-2}$. We first consider the upper bound to E_0 . Using the ground state Ψ_0 for $\varphi = 0$ as a trial function, we immediately obtain

$$E_0(N, a, \varphi) \leq \langle \Psi_0, H_N^A \Psi_0 \rangle = E_0(N, a, 0) + N\mu \frac{\varphi^2}{L^2}, \quad (14)$$

since $\langle \Psi_0, \nabla_j \Psi_0 \rangle = 0$. From Ref. 16 we know that $E_0(N, a, 0) \leq 4\pi\mu N \rho a [1 + (\text{const}) Y^{1/3}]$, which has the right form as $N \rightarrow \infty$.

For the lower bound to the ground state energy we need the following lemma. **LEMMA 1** (*Localization of energy*). For all symmetric, normalized wave functions $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ with periodic boundary conditions on $\mathcal{K}$, and for $N \geq Y^{-1/17}$,

$$\begin{aligned} \frac{1}{N} \langle \Psi, H_N^A \Psi \rangle \geq & [1 - (\text{const}) Y^{1/17}] \left(4\pi\mu\rho a \right. \\ & \left. + \mu \int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\Omega_{\mathbf{X}}} d\mathbf{r}_1 |(\nabla_1 + i\varphi L)\Psi(\mathbf{r}_1, \mathbf{X})|^2 \right), \end{aligned} \quad (15)$$

where $\mathbf{X} = (\mathbf{r}_2, \dots, \mathbf{r}_N)$, $d\mathbf{X} = \prod_{j=2}^N d\mathbf{r}_j$, and

$$\Omega_{\mathbf{X}} = \{\mathbf{r}_1 : \min_{j \geq 2} |\mathbf{r}_1 - \mathbf{r}_j| \geq R\} \quad (16)$$

with $R = aY^{-5/17}$.

Proof. Since Ψ is symmetric, the left side of Eq. (15) can be written as

$$\int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\mathcal{K}} d\mathbf{r}_1 \left[\mu |(\nabla_1 + i\varphi/L)\Psi(\mathbf{r}_1, \mathbf{X})|^2 + \frac{1}{2} \sum_{j=2}^N v(|\mathbf{r}_1 - \mathbf{r}_j|) |\Psi(\mathbf{r}_1, \mathbf{X})|^2 \right]. \quad (17)$$

For any $\varepsilon > 0$ and $R > 0$ this is

$$\geq \varepsilon T + (1 - \varepsilon)(T^{\text{in}} + I) + (1 - \varepsilon)T_{\varphi}^{\text{out}}, \quad (18)$$

with

$$T = \mu \int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\mathcal{K}} d\mathbf{r}_1 |\nabla_1 |\Psi(\mathbf{r}_1, \mathbf{X})||^2, \quad (19)$$

$$T^{\text{in}} = \mu \int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\Omega_{\mathbf{X}}^c} d\mathbf{r}_1 |\nabla_1 |\Psi(\mathbf{r}_1, \mathbf{X})||^2, \quad (20)$$

$$T_{\varphi}^{\text{out}} = \mu \int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\Omega_{\mathbf{X}}} d\mathbf{r}_1 |(\nabla_1 + i\varphi/L)\Psi(\mathbf{r}_1, \mathbf{X})|^2, \quad (21)$$

and

$$I = \frac{1}{2} \int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\mathcal{K}} d\mathbf{r}_1 \sum_{j=2}^N v(|\mathbf{r}_1 - \mathbf{r}_j|) |\Psi(\mathbf{r}_1, \mathbf{X})|^2. \quad (22)$$

Here

$$\Omega_{\mathbf{X}}^c = \{\mathbf{r}_1 : |\mathbf{r}_1 - \mathbf{r}_j| < R \text{ for some } j \geq 2\} \quad (23)$$

is the complement of $\Omega_{\mathbf{X}}$, and the diamagnetic inequality $|(\nabla + i\varphi/L)f(\mathbf{r})|^2 \geq |\nabla|f(\mathbf{r})||^2$ has been used. The proof is completed by using the results of Refs. 15 and 14 (also see Ref. 20) which tell us that for $\varepsilon = Y^{1/17}$ and $R = aY^{-5/17}$,

$$\varepsilon T + (1 - \varepsilon)(T^{\text{in}} + I) \geq [1 - (\text{const})Y^{1/17}]4\pi\mu\rho a \quad (24)$$

as long as $N \geq Y^{-1/17}$. (This estimate is highly nontrivial. Among several other things it uses a generalization of Dyson's lemma.¹⁹) Q.E.D.

The following lemma 2 is needed for a lower bound on the second term in Eq. (15). It is stated for $\mathcal{K}$ the $L \times L \times L$ -cube with periodic boundary conditions, but it can be generalized to arbitrary connected sets $\mathcal{K}$ that are sufficiently nice so that the Rellich-Kondrashov theorem [see Ref. 17 (Thm. 8.9)] holds on $\mathcal{K}$. In particular, this is the case if $\mathcal{K}$ has the ‘‘cone property.’’¹⁷ Another possible generalization is to include general bounded vector fields replacing φ ; see Ref. 21.

If Ω is any subset of $\mathcal{K}$ we shall denote $\int_{\Omega} f^*(\mathbf{r})g(\mathbf{r})d\mathbf{r}$ by $\langle f, g \rangle_{\Omega}$ and $\langle f, f \rangle_{\Omega}^{1/2}$ by $\|f\|_{L^2(\Omega)}$. We also denote $\nabla + i\varphi$ by ∇_{φ} for short.

LEMMA 2 (Generalized Poincaré inequality). *For any $|\varphi| < \pi$ there are constants $c > 0$ and $C < \infty$ such that for all subsets $\Omega \subset \mathcal{K}$ and all functions f on the torus $\mathcal{K}$ the following estimate holds:*

$$\begin{aligned} \|\nabla_{\varphi} f\|_{L^2(\Omega)}^2 &\geq \frac{\varphi^2}{L^2} \|f\|_{L^2(\mathcal{K})}^2 + \frac{c}{L^2} \|f - L^{-3}\langle 1, f \rangle_{\mathcal{K}}\|_{L^2(\mathcal{K})}^2 \\ &\quad - C \left(\|\nabla_{\varphi} f\|_{L^2(\mathcal{K})}^2 + \frac{1}{L^2} \|f\|_{L^2(\mathcal{K})}^2 \right) \left(\frac{|\Omega|^c}{|\mathcal{K}|} \right)^{1/2}. \end{aligned} \quad (25)$$

Here $|\Omega|^c$ is the volume of $\Omega^c = \mathcal{K} \setminus \Omega$, the complement of Ω in $\mathcal{K}$.

Proof. We shall derive Eq. (25) from a special form of this inequality that holds for all functions that are orthogonal to the constant function. That is, for any positive $\alpha < 2/3$ and some constants $c > 0$ and $\tilde{C} < \infty$ (depending only on α and $|\varphi| < \pi$) we claim that

$$\|\nabla_{\varphi} h\|_{L^2(\Omega)}^2 \geq \frac{\varphi^2 + c}{L^2} \|h\|_{L^2(\mathcal{K})}^2 - \tilde{C} \left(\frac{|\Omega|^c}{|\mathcal{K}|} \right)^{\alpha} \|\nabla_{\varphi} h\|_{L^2(\mathcal{K})}^2, \quad (26)$$

provided $\langle 1, h \rangle_{\mathcal{K}} = 0$. [Remark: Eq. (26) holds also for $\alpha = 2/3$, but the proof is slightly more complicated in that case. See Ref. 21] If (26) is known the derivation of Eq. (25) is easy: For any f , the function $h = f - L^{-3}\langle 1, f \rangle_{\mathcal{K}}$ is orthogonal to 1. Moreover,

$$\begin{aligned} \|\nabla_{\varphi} h\|_{L^2(\Omega)}^2 &= \|\nabla_{\varphi} h\|_{L^2(\mathcal{K})}^2 - \|\nabla_{\varphi} h\|_{L^2(\Omega^c)}^2 \\ &= \|\nabla_{\varphi} f\|_{L^2(\Omega)}^2 - \frac{\varphi^2}{L^2} |\langle L^{-3/2}, f \rangle_{\mathcal{K}}|^2 \left(1 + \frac{|\Omega^c|}{|\mathcal{K}|} \right) \\ &\quad + 2 \frac{\varphi}{L} \text{Re} \langle L^{-3/2}, f \rangle_{\mathcal{K}} \langle \nabla_{\varphi} f, L^{-3/2} \rangle_{\Omega^c} \\ &\leq \|\nabla_{\varphi} f\|_{L^2(\Omega)}^2 - \frac{\varphi^2}{L^2} |\langle L^{-3/2}, f \rangle_{\mathcal{K}}|^2 \\ &\quad + \frac{|\varphi|}{L} \left(L \|\nabla_{\varphi} f\|_{L^2(\mathcal{K})}^2 + \frac{1}{L} \|f\|_{L^2(\mathcal{K})}^2 \right) \left(\frac{|\Omega^c|}{|\mathcal{K}|} \right)^{1/2} \end{aligned} \quad (27)$$

and

$$\begin{aligned} \frac{\varphi^2 + c}{L^2} \|h\|_{L^2(\mathcal{K})}^2 &= \frac{\varphi^2}{L^2} (\|f\|_{L^2(\mathcal{K})}^2 - |\langle L^{-3/2}, f \rangle_{\mathcal{K}}|^2) \\ &\quad + \frac{c}{L^2} \|f - L^{-3}\langle 1, f \rangle_{\mathcal{K}}\|_{L^2(\mathcal{K})}^2. \end{aligned} \quad (28)$$

Setting $\alpha = \frac{1}{2}$, using $\|\nabla_{\varphi} h\|_{L^2(\mathcal{K})} \leq \|\nabla_{\varphi} f\|_{L^2(\mathcal{K})}$ in the last term in Eq. (26) and combining Eqs. (26), (27), and (28) gives Eq. (25) with $C = |\varphi| + \tilde{C}$.

We now turn to the proof of Eq. (26). For simplicity we set $L = 1$. The general case follows by scaling. Assume that Eq. (26) is false. Then there exist sequences of constants $C_n \rightarrow \infty$, functions h_n with $\|h_n\|_{L^2(\mathcal{K})} = 1$ and $\langle 1, h_n \rangle_{\mathcal{K}} = 0$, and domains $\Omega_n \subset \mathcal{K}$ such that

$$\lim_{n \rightarrow \infty} \left(\|\nabla_{\varphi} h_n\|_{L^2(\Omega_n)}^2 + C_n |\Omega_n^c|^{\alpha} \|\nabla_{\varphi} h_n\|_{L^2(\mathcal{K})}^2 \right) \leq \varphi^2. \quad (29)$$

We shall show that this leads to a contradiction.

Since the sequence h_n is bounded in $L^2(\mathcal{K})$ it has a subsequence, denoted again by h_n , that converges weakly to

SUPERFLUIDITY IN DILUTE TRAPPED BOSE GASES

PHYSICAL REVIEW B 66, 134529 (2002)

some $h \in L^2(\mathcal{K})$ [i.e., $\langle g, h_n \rangle_{\mathcal{K}} \rightarrow \langle g, h \rangle_{\mathcal{K}}$ for all $g \in L^2(\mathcal{K})$]. Moreover, by Hölder's inequality the $L^p(\Omega_n^c)$ norm $\|\nabla_\varphi h_n\|_{L^p(\Omega_n^c)} = (\int_{\Omega_n^c} |h(\mathbf{r})|^p d\mathbf{r})^{1/p}$ is bounded by $|\Omega_n^c|^{\alpha/2} \|\nabla_\varphi h_n\|_{L^2(\mathcal{K})}$ for $p = 2/(\alpha + 1)$. From Eq. (29) we conclude that $\|\nabla_\varphi h_n\|_{L^p(\Omega_n^c)}$ is bounded and also that $\|\nabla_\varphi h_n\|_{L^p(\Omega_n)} \leq \|\nabla_\varphi h_n\|_{L^2(\Omega_n)}$ is bounded. Altogether, $\nabla_\varphi h_n$ is bounded in $L^p(\mathcal{K})$, and by passing to a further subsequence if necessary, we can therefore assume that $\nabla_\varphi h_n$ converges weakly in $L^p(\mathcal{K})$. The same applies to ∇h_n . Since $p = 2/(\alpha + 1)$ with $\alpha < 2/3$ the hypotheses of the Rellich-Kondrashov Theorem [see Ref. 17 (Thm 8.9)] are fulfilled and consequently h_n converges strongly in $L^2(\mathcal{K})$ to h (i.e., $\|h - h_n\|_{L^2(\mathcal{K})} \rightarrow 0$). We shall now show that

$$\liminf_{n \rightarrow \infty} \|\nabla_\varphi h_n\|_{L^2(\Omega_n)}^2 \geq \|\nabla_\varphi h\|_{L^2(\mathcal{K})}^2. \quad (30)$$

This will complete the proof because the h_n are normalized and orthogonal to 1 and the same holds for h by strong convergence. Hence the right side of Eq. (30) is necessarily $> \varphi^2$, since for $|\varphi| < \pi$ the lowest eigenvalue of $-\nabla_\varphi^2$, with constant eigenfunction, is non-degenerate. This contradicts Eq. (29).

Equation (30) is essentially a consequence of the weak lower semicontinuity of the L^2 norm, but the dependence on Ω_n leads to a slight complication. First, Eq. (29) and $C_n \rightarrow \infty$ clearly imply that $|\Omega_n^c| \rightarrow 0$, because $\|\nabla_\varphi h_n\|_{L^2(\mathcal{K})}^2 > \varphi^2$. By choosing a subsequence we may assume that $\sum_n |\Omega_n^c| < \infty$. For some fixed N let $\tilde{\Omega}_N = \mathcal{K} \setminus \bigcup_{n \geq N} \Omega_n^c$. Then $\tilde{\Omega}_N \subset \Omega_n$ for $n \geq N$. Since $\|\nabla_\varphi h_n\|_{L^2(\Omega_n)}$ is bounded, $\nabla_\varphi h_n$ is also bounded in $L^2(\tilde{\Omega}_N)$ and a subsequence of it converges weakly in $L^2(\tilde{\Omega}_N)$ to $\nabla_\varphi h$. Hence

$$\begin{aligned} \liminf_{n \rightarrow \infty} \|\nabla_\varphi h_n\|_{L^2(\Omega_n)}^2 &\geq \liminf_{n \rightarrow \infty} \|\nabla_\varphi h_n\|_{L^2(\tilde{\Omega}_N)}^2 \\ &\geq \|\nabla_\varphi h\|_{L^2(\tilde{\Omega}_N)}^2. \end{aligned} \quad (31)$$

Since $\tilde{\Omega}_N \subset \tilde{\Omega}_{N+1}$ and $\bigcup_N \tilde{\Omega}_N = \mathcal{K}$ (up to a set of measure zero), we can now let $N \rightarrow \infty$ on the right side of Eq. (31). By monotone convergence this converges to $\|\nabla_\varphi h\|_{L^2(\mathcal{K})}^2$. This proves Eq. (30) which, as remarked above, contradicts Eq. (29). Q.E.D.

We now are able to finish the proof of theorem 1. From lemmas 1 and 2 we infer that, for any symmetric Ψ with $\langle \Psi, \Psi \rangle = 1$ and for N large enough,

$$\begin{aligned} &\frac{1}{N} \langle \Psi, H'_N \Psi \rangle [1 - (\text{const}) Y^{1/17}]^{-1} \\ &\geq 4\pi\mu\rho a + \mu \frac{\varphi^2}{L^2} \\ &\quad - CY^{1/17} \left(\frac{1}{L^2} + \frac{1}{N} \left\langle \Psi, \sum_j (\nabla_j + i\varphi) \Psi \right\rangle \right) \\ &\quad + \frac{c}{L^2} \int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\mathcal{K}} d\mathbf{r}_1 \left| \Psi(\mathbf{r}_1, \mathbf{X}) \right. \\ &\quad \left. - L^{-3} \left[\int_{\mathcal{K}} d\mathbf{r} \Psi_0(\mathbf{r}, \mathbf{X}) \right] \right|^2, \end{aligned} \quad (32)$$

where we used that $|\Omega^c| \leq (4\pi/3)NR^3 = (\text{const})L^3Y^{2/17}$. From this we can infer two things. First, since the kinetic energy, divided by N , is certainly bounded independent of N , as the upper bound shows, we obtain that

$$\liminf_{N \rightarrow \infty} \frac{E_0(N, a, \varphi)}{N} \geq 4\pi\mu\rho a + \mu \frac{\varphi^2}{L^2} \quad (33)$$

for any $|\varphi| < \pi$. By continuity this holds also for $|\varphi| = \pi$, proving Eq. (6). (To be precise, $E_0/N - \mu\varphi^2L^{-2}$ is concave in φ , and therefore stays concave, and in particular continuous, in the limit $N \rightarrow \infty$.) Second, since the upper and the lower bound to E_0 agree in the limit considered, the positive last term in Eq. (32) has to vanish in the limit. That is, we obtain that for the ground state wave function Ψ_0 of H'_N ,

$$\begin{aligned} &\lim_{N \rightarrow \infty} \int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\mathcal{K}} d\mathbf{r}_1 \left| \Psi_0(\mathbf{r}_1, \mathbf{X}) \right. \\ &\quad \left. - L^{-3} \left[\int_{\mathcal{K}} d\mathbf{r} \Psi_0(\mathbf{r}, \mathbf{X}) \right] \right|^2 = 0. \end{aligned} \quad (34)$$

This proves Eq. (7), since

$$\begin{aligned} &\int_{\mathcal{K}^{N-1}} d\mathbf{X} \int_{\mathcal{K}} d\mathbf{r}_1 \left| \Psi_0(\mathbf{r}_1, \mathbf{X}) - L^{-3} \left[\int_{\mathcal{K}} d\mathbf{r} \Psi_0(\mathbf{r}, \mathbf{X}) \right] \right|^2 \\ &= 1 - \frac{1}{NL^3} \int_{\mathcal{K} \times \mathcal{K}} \gamma(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}', \end{aligned} \quad (35)$$

and therefore $N^{-1} \langle L^{-3/2} \gamma_N | L^{-3/2} \rangle \rightarrow 1$. As explained in Refs. 14 and 20 this suffices for the convergence $N^{-1} \gamma_N \rightarrow |L^{-3/2} \rangle \langle L^{-3/2}|$ in trace class norm. Q.E.D.

IV. CONCLUSIONS

We have shown that a Bose gas with short range, repulsive interactions is both a 100% superfluid and also 100% Bose-Einstein condensed in its ground state in the Gross-Pitaevskii limit where the parameter Na/L is kept fixed as $N \rightarrow \infty$. This is a simultaneous large N and low density limit, because the dimensionless density parameter ρa^3 is here proportional to $1/N^2$. If ρa^3 is not zero, but small, a depletion of the Bose-Einstein condensate of the order $(\rho a^3)^{1/2}$ is expected (see, e.g., Ref. 22). Nevertheless, complete superfluidity in the ground state, e.g., of Helium 4, is experimentally observed. It is an interesting open problem to deduce this property rigorously from first principles. In the case of a one-dimensional hard-core Bose gas superfluidity in the ground state is easy to show, but nevertheless there is no Bose-Einstein condensation at all, not even in the ground state.^{9,10}

ACKNOWLEDGMENTS

E.H.L. was partially supported by the U.S. National Science Foundation, Grant No. PHY 98-20650. R.S. was supported by the Austrian Science Fund in the form of an Erwin Schrödinger fellowship.

ELLIOTT H. LIEB, ROBERT SEIRINGER, AND JAKOB YNGVASON

PHYSICAL REVIEW B 66, 134529 (2002)

*Electronic address: lieb@math.princeton.edu

†On leave from Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, A-1090 Vienna, Austria; Electronic address: rseiringer@math.princeton.edu

‡Electronic address: yngvason@thor.thp.univie.ac.at

¹D.R. Tilley and J. Tilley, *Superfluidity and Superconductivity*, 3rd ed. (Hilger, Bristol, 1990).

²P.C. Hohenberg and P.C. Martin, Ann. Phys. (N.Y.) **34**, 291 (1965).

³The ground state with $v=0$ remains an eigenstate of the Hamiltonian with arbitrary v since its total momentum is zero. Its energy is $\frac{1}{2}mNv^2$ above the ground state energy for $v=0$. Since in a finite box the spectrum of the Hamiltonian for arbitrary v is discrete and the energy gap above the ground state is bounded away from zero for v small, the ground state for $v=0$ is at the same time the ground state of the Hamiltonian with v if $\frac{1}{2}mNv^2$ is smaller than the gap.

⁴N.V. Prokof'ev and B.V. Svistunov, Phys. Rev. B **61**, 11 282 (2000).

⁵K. Huang, in *Bose-Einstein Condensation*, edited by A. Griffin, D.W. Stroke, and S. Stringari (Cambridge University Press, Cambridge, 1995), pp. 31–50.

⁶G.E. Astrakharchik, J. Boronat, J. Casulleras, and S. Giorgini, cond-mat/0111165 (unpublished).

⁷M. Kobayashi and M. Tsubota, cond-mat/0202364 (unpublished).

⁸M. Girardeau, J. Math. Phys. **1**, 516 (1960).

⁹A. Lenard, J. Math. Phys. **5**, 930 (1964).

¹⁰L. Pitaevskii and S. Stringari, J. Low Temp. Phys. **85**, 377 (1991).

¹¹G. Baym, in *Math. Methods in Solid State and Superfluid Theory*, Scottish University Summer School of Physics (Oliver and Boyd, Edinburgh, 1969).

¹²A.J. Leggett, Rev. Mod. Phys. **71**, S318 (1999).

¹³F. Dalfovo, S. Giorgini, L. Pitaevskii, and S. Stringari, Rev. Mod. Phys. **71**, 463 (1999).

¹⁴E.H. Lieb and R. Seiringer, Phys. Rev. Lett. **88**, 170409 (2002).

¹⁵E.H. Lieb and J. Yngvason, Phys. Rev. Lett. **80**, 2504 (1998).

¹⁶E.H. Lieb, R. Seiringer, and J. Yngvason, Phys. Rev. A **61**, 043602 (2000).

¹⁷E.H. Lieb and M. Loss, *Analysis*, 2nd ed. (American Mathematical Society, Providence, 2001).

¹⁸A.L. Fetter and A.A. Svidzinsky, J. Phys.: Condens. Matter **13**, R135 (2001).

¹⁹F.J. Dyson, Phys. Rev. **106**, 317 (1957).

²⁰E.H. Lieb, R. Seiringer, J.P. Solovej, and J. Yngvason, *Contemporary Developments in Mathematics 2001* (International Press Boston, in press) [math-ph/0204027 (unpublished)].

²¹E.H. Lieb, R. Seiringer, and J. Yngvason, math.FA/0205088 (unpublished).

²²C.J. Pethick and H. Smith, *Bose-Einstein Condensation in Dilute Gases* (Cambridge University Press, Cambridge, 2002).

One-Dimensional Bosons in Three-Dimensional Traps

Elliott H. Lieb and Robert Seiringer*

Department of Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton, New Jersey 08544, USA

Jakob Yngvason

Institut für Theoretische Physik, Universität Wien, Boltzmannngasse 5, A-1090 Vienna, Austria

(Received 2 April 2003; published 7 October 2003)

Recent experimental and theoretical work has indicated conditions in which a trapped, low density Bose gas ought to behave like the 1D delta-function Bose gas solved by Lieb and Liniger. Up until now, the theoretical arguments have been based on variational/perturbative ideas or numerical investigations. There are four parameters: density, transverse and longitudinal dimensions, and scattering length. In this paper we explicate five parameter regions in which various types of 1D or 3D behavior occur in the ground state. Our treatment is based on a rigorous analysis of the many-body Schrödinger equation.

DOI: 10.1103/PhysRevLett.91.150401

PACS numbers: 05.30.Jp, 03.75.Hh, 67.40.-w

It appears to be possible to do experiments in highly elongated traps on ultracold Bose gases that are effectively 1D. More precisely, the 1D Bose gas with a delta-function two-body interaction, analyzed long ago [1,2], should be visible, with its quasifermionic behavior [3], the absence of Bose-Einstein condensation (BEC) in a dilute limit [4], and an excitation spectrum different from that predicted by Bogoliubov's theory [2,5]. Several theoretical investigations on the transitions from 3D to an effective 1D behavior were triggered by [6]. See, e.g., [7–9]. Systems showing indications of such a transition have recently been prepared experimentally [10].

The theoretical work on the dimensional crossover for the ground state in elongated traps has so far been based either on variational calculations, starting from a 3D delta potential [6,9], or on numerical quantum Monte Carlo studies [11] with more realistic, genuine 3D potentials, but particle numbers limited to the order of 100. This work is important and has led to valuable insights, in particular about different parameter regions [7], but a more thorough theoretical understanding is clearly desirable since this is not a simple problem. In fact, it is evident that for a potential with a hard core the true 3D wave functions do not approximately factorize in the longitudinal and transverse variable (otherwise the energy would be infinite) and the effective 1D potential cannot be obtained by simply integrating out the transverse variables of the 3D potential (that would immediately create an impenetrable barrier in 1D). It is important to be able to demonstrate rigorously, and therefore unambiguously, that the 1D behavior really follows from the fundamental Schrödinger equation. It is also important to delineate, as we do here, precisely what can be seen in the different parameter regions. The full proofs of our assertions are long and will be given elsewhere [12], but we emphasize that everything can be rigorously derived from first principles. In this Letter we state our main results and outline the basic ideas for the proofs.

We shall always be concerned with the ground state and with large particle number, $N \gg 1$, which is appropriate for the consideration of actual experiments. Other parameters of the problem are the scattering length, a , of the two-body interaction potential and two lengths, r and L , describing the transverse and the longitudinal extension of the trap potential, respectively.

It is convenient to write the Hamiltonian in the following way (in units where $\hbar = 2m = 1$):

$$H_{N,L,r,a} = \sum_{j=1}^N [-\nabla_j^2 + V_r^\perp(\mathbf{x}_j^\perp) + V_L(z_j)] + \sum_{1 \leq i < j \leq N} v_a(|\mathbf{x}_i - \mathbf{x}_j|), \quad (1)$$

with $\mathbf{x} = (x, y, z) = (\mathbf{x}^\perp, z)$ and with

$$V_r^\perp(\mathbf{x}^\perp) = r^{-2} V^\perp(\mathbf{x}^\perp/r), \quad V_L(z) = L^{-2} V(z/L), \\ v_a(|\mathbf{x}|) = a^{-2} v(|\mathbf{x}|/a). \quad (2)$$

Here, r, L, a are variable scaling parameters while $V^\perp, V$, and v are fixed. The interaction potential v is supposed to be non-negative, of finite range, and have scattering length 1; the scaled potential v_a then has scattering length a . The external trap potentials V and $V^\perp$ confine the motion in the longitudinal (z) and the transversal ($\mathbf{x}^\perp$) directions, respectively, and are assumed to be continuous and tend to ∞ as $|z|$ and $|\mathbf{x}^\perp|$ tend to ∞ . To simplify the discussion we find it also convenient to assume that V is homogeneous of some order $s > 0$, namely $V(z) = |z|^s$, but weaker assumptions, e.g., asymptotic homogeneity [13], would in fact suffice. The case of a simple box with hard walls is realized by taking $s = \infty$, while the usual harmonic approximation is $s = 2$. It is understood that the lengths associated with the ground states of $-d^2/dz^2 + V(z)$ and $-(\nabla^\perp)^2 + V^\perp(\mathbf{x}^\perp)$ are both of the order of 1 so that L and r measure, respectively, the longitudinal and the transverse extensions of the trap.

We denote the ground state energy of (1) by $E^{\text{QM}}(N, L, r, a)$ and the ground state particle density by $\rho_{N,L,r,a}^{\text{QM}}(\mathbf{x})$.

In parallel with the 3D Hamiltonian we consider the Hamiltonian for n bosons in 1D with delta interaction and coupling constant $g \geq 0$ (denoted $2c$ in [1]); i.e.,

$$H_{n,g}^{\text{1D}} = \sum_{j=1}^n -\partial^2/\partial z_j^2 + g \sum_{1 \leq i < j \leq n} \delta(z_i - z_j). \quad (3)$$

We consider the Hamiltonian for the z_j in an interval of length ℓ in the thermodynamic limit, $\ell \rightarrow \infty$, $n \rightarrow \infty$, with $\rho = n/\ell$ fixed. The ground state energy per particle in this limit is independent of boundary conditions and can, according to Lieb and Liniger [1], be written as

$$e_0^{\text{1D}}(\rho) = \rho^2 e(g/\rho), \quad (4)$$

with a function $e(t)$ determined by a certain integral equation. Its asymptotic form is $e(t) \approx \frac{1}{2}t$ for $t \ll 1$ and $e(t) \rightarrow \pi^2/3$ for $t \rightarrow \infty$. Thus

$$e_0^{\text{1D}}(\rho) \approx \frac{1}{2}g\rho \quad \text{for } g/\rho \ll 1, \quad (5)$$

$$e_0^{\text{1D}}(\rho) \approx (\pi^2/3)\rho^2 \quad \text{for } g/\rho \gg 1. \quad (6)$$

These two situations correspond to high 1D density (weak interaction) and low 1D density (strong interaction), respectively. The latter case is usually referred to as the Girardeau-Tonks regime. Physically, the main difference is that in the strong interaction regime the motion of the particles in the longitudinal direction is highly correlated, while in the weak interaction regime it is not.

Taking $\rho e_0^{\text{1D}}(\rho)$ as a local energy density for an inhomogeneous 1D system we can form the energy functional

$$\mathcal{E}[\rho] = \int_{-\infty}^{\infty} [|\nabla \sqrt{\rho}(z)|^2 + V_L(z)\rho(z) + \rho(z)^3 e(g/\rho(z))] dz. \quad (7)$$

The gradient term represents additional kinetic energy associated with the inhomogeneity of the gas that is not accounted for by the “local” kinetic energy included in the last term. The ground state energy of this functional is defined to be $E^{\text{1D}}(N, L, g) = \inf\{\mathcal{E}[\rho]; \rho(z) \geq 0, \int \rho(z) dz = N\}$. By standard methods (cf., e.g., [14]) one can show that there is a unique minimizer, i.e., a density $\rho_{N,L,g}(z)$ with $\int \rho_{N,L,g}(z) dz = N$ and $\mathcal{E}[\rho_{N,L,g}] = E^{\text{1D}}(N, L, g)$. We define the mean 1D density of this minimizer to be $\bar{\rho} = N^{-1} \int_{-\infty}^{\infty} (\rho_{N,L,g}(z))^2 dz$. In the rigid box, i.e., for $s = \infty$, $\bar{\rho}$ is simply N/L (except for boundary corrections), but in more general traps it depends also on g besides N and L . The order of magnitude of $\bar{\rho}$ in various regions of the parameters will be described below.

Our main result relates the 3D ground state energy of (1), $E^{\text{QM}}(N, L, r, a)$, to the 1D density functional energy $E^{\text{1D}}(N, L, g)$ in the large N limit with $g \sim a/r^2$ provided r/L and a/r are sufficiently small. To state this precisely,

150401-2

let $e^\perp$ and $b(\mathbf{x}^\perp)$, respectively, denote the ground state energy and the normalized ground state wave function of $-(\nabla^\perp)^2 + V^\perp(\mathbf{x}^\perp)$. The corresponding quantities for $-(\nabla^\perp)^2 + V_r^\perp(\mathbf{x}^\perp)$ are $e^\perp/r^2$ and $b_r(\mathbf{x}^\perp) = (1/r)b(\mathbf{x}^\perp/r)$. In the case that the trap is a cylinder with hard walls b is a Bessel function; for a quadratic $V^\perp$ it is a Gaussian.

Define g by

$$g = \frac{8\pi a}{r^2} \int |b(\mathbf{x}^\perp)|^4 d^2 \mathbf{x}^\perp = 8\pi a \int |b_r(\mathbf{x}^\perp)|^4 d^2 \mathbf{x}^\perp. \quad (8)$$

THEOREM: Let $N \rightarrow \infty$ and simultaneously $r/L \rightarrow 0$ and $a/r \rightarrow 0$ in such a way that $r^2 \bar{\rho} \cdot \min\{\bar{\rho}, g\} \rightarrow 0$. Then

$$\lim \frac{E^{\text{QM}}(N, L, r, a) - N e^\perp/r^2}{E^{\text{1D}}(N, L, g)} = 1. \quad (9)$$

Moreover, if we define the 1D quantum-mechanical density by averaging over the transverse variables, i.e.,

$$\bar{\rho}_{N,L,r,a}^{\text{QM}}(z) \equiv \int \rho_{N,L,r,a}^{\text{QM}}(\mathbf{x}^\perp, z) d^2 \mathbf{x}^\perp, \quad (10)$$

then $\bar{\rho}_{N,L,r,a}^{\text{QM}}(z)/\rho_{N,L,g}(z) \rightarrow 1$ in a suitable sense.

Note that because of (5) and (6) the condition $r^2 \bar{\rho} \cdot \min\{\bar{\rho}, g\} \rightarrow 0$ is the same as $e_0^{\text{1D}}(\bar{\rho}) \ll 1/r^2$, i.e., the average energy per particle associated with the longitudinal motion should be much smaller than the energy gap between the ground and first excited state of the confining Hamiltonian in the transverse directions. Thus, the basic physics is highly quantum mechanical and has no classical counterpart. The system can be described by a 1D functional (7), even though the transverse trap dimension is much larger than the range of the atomic forces.

The domain of the parameters can be divided into subregions characterized by specific restrictions on the size of the ratio $g/\bar{\rho}$ as $N \rightarrow \infty$. It turns out that there are five regions altogether, each described by a limiting case of the general density functional (7), but there is a basic dichotomy between the regions that can be regarded as limits of 3D Gross-Pitaevskii theory and those that cannot be reached in that way. The former (regions 1–3 below) are characterized by the condition that the 3D ground state energy per particle, which is proportional to a times the three dimensional density $\rho^{3D} \sim N/(r^2 L)$ for a dilute gas [15], is much smaller than the energy given by the formula (6). This means that $g/\bar{\rho} \ll 1$ with g given by (8). In the latter regime (regions 4 and 5 below) the energy (6) is comparable to or smaller than the 3D energy. In both regimes the internal energy of the gas is small compared to the energy of transversal confinement which is of the order N/r^2 . However, this in itself does not imply a specifically 1D behavior. (If a is sufficiently small it is satisfied in a trap of any shape.) 1D behavior, when it occurs, manifests itself by the fact that the transverse motion of the atoms is uncorrelated while the longitudinal motion is correlated (very roughly speaking) in the

150401-2

same way as pearls on a necklace. Thus, the true criterion for 1D behavior is that $g/\bar{\rho}$ is of the order unity or larger, and not merely the condition that the energy of confinement dominates the internal energy.

We shall now briefly describe the division of the two regimes into the five subregions. We always assume $N \rightarrow \infty$ and $r/L \rightarrow 0$.

REGION 1: The ideal gas case.— $g/\bar{\rho} \ll N^{-2}$, with $\bar{\rho} \sim N/L$, corresponding to the trivial case where the interaction is so weak that it effectively vanishes in the large N limit and everything collapses to the ground state of $-d^2/dz^2 + V(z)$ with ground state energy e^{ll} . The energy E^{1D} in (9) can be replaced by $N e^{\text{ll}}/L^2$. Note that $g/\bar{\rho} \ll N^{-2}$ means that the 3D interaction energy $\sim \rho^{3\text{D}} a \ll 1/L^2$.

REGION 2: The 1D GP case.— $g/\bar{\rho} \sim N^{-2}$, with $\bar{\rho} \sim N/L$, described by a 1D Gross-Pitaevskii energy functional

$$\mathcal{E}^{\text{GP}}[\rho] = \int_{-\infty}^{\infty} [|\nabla \sqrt{\rho}(z)|^2 + V_L(z)\rho(z) + \frac{1}{2}g\rho(z)^2]dz, \quad (11)$$

corresponding to the high density approximation (5) of the interaction energy in (7). Its ground state energy, $E^{\text{GP}} = \inf\{\mathcal{E}^{\text{GP}}[\rho]; \rho(z) \geq 0, \int \rho = N\}$, has the scaling property $E^{\text{GP}}(N, L, g) = NL^{-2}E^{\text{GP}}(1, 1, NgL)$.

REGION 3: The 1D TF case.— $N^{-2} \ll g/\bar{\rho} \ll 1$, with $\bar{\rho} \sim (N/L)(NgL)^{-1/(s+1)}$, where s is the degree of homogeneity of the longitudinal confining potential V . This region is described by a Thomas-Fermi type functional

$$\mathcal{E}^{\text{TF}}[\rho] = \int_{-\infty}^{\infty} [V_L(z)\rho(z) + \frac{1}{2}g\rho(z)^2]dz. \quad (12)$$

It is a limiting case of region 2 in the sense that $NgL \sim NaL/r^2 \rightarrow \infty$, but a/r is sufficiently small so that $g/\bar{\rho} \sim (aL/Nr^2)(NaL/r^2)^{1/(s+1)} \rightarrow 0$; i.e., the high density approximation in (5) is still valid. In this limit the gradient term in (11) becomes vanishingly small compared to the other terms.

REGION 4: The LL case.— $g/\bar{\rho} \sim 1$, with $\bar{\rho} \sim (N/L)N^{-2/(s+2)}$, described by an energy functional

$$\mathcal{E}^{\text{LL}}[\rho] = \int_{-\infty}^{\infty} [V_L(z)\rho(z) + \rho(z)^3 e(g/\rho(z))]dz. \quad (13)$$

This region corresponds to the case $g/\bar{\rho} \sim 1$, so that neither the high density (5) nor the low density approximation (6) is valid and the full LL energy (4) has to be used, but, as in region 3, the gradient term in (7) is negligible. The scaling of the ground state energy of (13) is $E^{\text{LL}}(N, L, g) = N\gamma^2 E^{\text{LL}}(1, 1, g/\gamma)$ with $\gamma = (N/L)N^{-2/(s+2)}$.

REGION 5: The GT case.— $g/\bar{\rho} \gg 1$, with $\bar{\rho} \sim (N/L)N^{-2/(s+2)}$, described by a functional with energy density $\sim \rho^3$, corresponding to the Girardeau-Tonks limit of the LL energy density. It corresponds to impenetrable particles, i.e., the limiting case $g/\bar{\rho} \rightarrow \infty$ and hence for-

mula (6) for the energy density. The energy functional is

$$\mathcal{E}^{\text{GT}}[\rho] = \int_{-\infty}^{\infty} [V_L(z)\rho(z) + (\pi^2/3)\rho(z)^3]dz, \quad (14)$$

with minimum energy $E^{\text{GT}}(N, L) = N\gamma^2 E^{\text{GT}}(1, 1)$.

We note that the condition $g/\bar{\rho} \sim 1$ means that region 4 requires the gas cloud to have aspect ratio r/L of the order $N^{-1}(a/r)$ or smaller, where $\bar{L} \sim LN^{2/(s+2)}$ is the length of the cloud. Experimentally, such small aspect ratios are quite a challenge and the situations described in [10] are still rather far from this regime. It may not be completely out of reach, however.

Regions 1–3 can be reached as limiting cases of a 3D Gross-Pitaevskii theory [12]. In this sense, the behavior in these regions contains remnants of the 3D theory, which also shows up in the fact that the proof of BEC in the 3D Gross-Pitaevskii limit in [16] can be carried over to regions 1 and 2 [12]. Heuristically, these traces of 3D can be understood from the fact that in regions 1–3 the 1D formula (5) for energy per particle, $g\rho \sim aN/(r^2L)$, gives the same result as the 3D formula [15], i.e., $a\rho^{3\text{D}}$. This is no longer so in regions 4 and 5.

We now comment on the main steps in the proof of the theorem, referring to [12] for full details. The different parameter regions have to be treated by different methods, a watershed lying between regions 1–3 on the one hand and regions 4–5 on the other. In regions 1–3, similar methods as in the proof of the 3D Gross-Pitaevskii limit theorem in [14] can be used. This 3D proof needs considerable modifications, however, because in [14] the external potential is fixed and the estimates are not uniform in the ratio r/L .

To prove (9) one has to establish upper and lower bounds, with controlled errors, on the QM many-body energy in terms of the energies obtained by minimizing the energy functionals appropriate for the various regions. The limit theorem for the densities can be derived from the energy estimates in a standard way by variation with respect to the external potential V_L . As usual, the upper bounds for the energy are easier than the lower bounds, but nevertheless not simple, in particular, not for “hard” potentials v .

The upper bound in regions 1–3 is obtained from a variational ansatz of the form $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = F(\mathbf{x}_1, \dots, \mathbf{x}_N) \prod_{k=1}^N b_r(\mathbf{x}_k) \sqrt{\rho^{\text{GP}}(z_k)}$, with $F(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{k=1}^N f(\mathbf{x}_k - \mathbf{x}_{j(k)})$, where $\mathbf{x}_{j(k)}$ is the nearest neighbor of $\mathbf{x}_k$ among the points $\mathbf{x}_j$, $j < k$, $b_r(\mathbf{x}^\perp)$ is the lowest eigenfunction of $-(\nabla^\perp)^2 + V_\perp^\perp$, and $\rho^{\text{GP}}(z)$ the minimizer of the 1D GP functional (11). The function f is, up to a cutoff length that has to be chosen optimally, the zero energy scattering solution for the two-body Hamiltonian with interaction V_a . The form of F , inspired by [17], is chosen rather than a Jastrow ansatz $\prod_{i<j} f(\mathbf{x}_i - \mathbf{x}_j)$, because it is computationally simpler for the purposes of obtaining rigorous estimates.

For an upper bound in regions 4–5 a natural variational ansatz would appear to be $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = F(\mathbf{x}_1, \dots, \mathbf{x}_N) \prod_{k=1}^N b_r(\mathbf{x}_k^{\perp}) \psi(z_1, \dots, z_N)$, where ψ is the ground state of $H_{N,g}^{\text{D}}$ with the external potential V_L added. However, in order to make a link with the exact solution (4) for a homogeneous gas, but also to control the norm of the trial function, it turns out to be necessary to localize the particles by dividing the trap into finite “boxes” (finite in z direction), with a finite particle number in each box and making the ansatz with the boundary condition $\Psi = 0$ for each box individually. The particles are then distributed optimally among the boxes to minimize the energy. This box method, but with the boundary condition $\nabla\Psi = 0$, is also used for the lower bounds to the energy. Another essential device for the lower bounds is *Dyson’s lemma* that was also used in [14–16]. This lemma, which goes back to Dyson’s seminal paper [17] on the hard-core Bose gas, estimates the kinetic and potential energy for a Hamiltonian with a hard potential v of finite range from below by the potential energy of a “soft” potential U of larger range but essentially the same scattering length as v . Borrowing a tiny part of the kinetic energy it is then possible to do perturbation theory with the soft potential U and use Temple’s inequality [18] to bound the errors. A direct application of perturbation theory to the original potential v , on the other hand, is in general not possible.

A core lemma for regions 4–5 is a lower bound on the 3D ground state energy in a finite box in terms of the 1D energy of the Hamiltonian (3) both with the boundary condition $\nabla\Psi = 0$. Denoting the former energy by $E_{\text{box}}^{\text{3D}}$ and the latter by $E_{\text{box}}^{\text{1D}}$, this bound for n particles in a box of length ℓ in the z direction reads

$$E_{\text{box}}^{\text{3D}} - \frac{ne^{\perp}}{r^2} \geq E_{\text{box}}^{\text{1D}} \left\{ 1 - Cn \left(\frac{a}{r} \right)^{1/8} \left[1 + \frac{nr}{\ell} \left(\frac{a}{r} \right)^{1/8} \right] \right\},$$

with a constant C . To prove this bound the ground state wave function is first written as a product of $\prod_k b_r(\mathbf{x}_k^{\perp})$ and a function $G(\mathbf{x}_1, \dots, \mathbf{x}_n)$. This subtracts $ne^{\perp}/r^2$ from $E_{\text{box}}^{\text{3D}}$ but the resulting minimization problem for G involves the weighted measure $\prod_k b_r(\mathbf{x}_k^{\perp})^2 d^3\mathbf{x}_k$ in place of $\prod_k d^3\mathbf{x}_k$. Nevertheless, Dyson’s lemma can be used, and after the hard potential v has been replaced by the soft potential U , it is possible to integrate the transverse variables away and obtain a minimization problem for a 1D many-body Hamiltonian with interaction $d(z_i - z_j) = a \iint b_r(\mathbf{x}_i^{\perp})^2 b_r(\mathbf{x}_j^{\perp})^2 U(\mathbf{x}_i - \mathbf{x}_j) d^2\mathbf{x}_i^{\perp} d^2\mathbf{x}_j^{\perp}$. In the limit considered this converges to a delta interaction with the coupling constant (8). The error terms in the estimate for $E_{\text{box}}^{\text{3D}}$ arise both from Temple’s inequality and the replacement of d by a delta function, among other things. When the particles are distributed optimally among the boxes to obtain a global lower bound, super-additivity of the energy and convexity of the energy density $\rho^3 e(g/\rho)$ are used, generalizing corresponding arguments in [15].

In conclusion, we have reported a rigorous analysis of the parameter regions in which a Bose gas in an elongated trap may or may not be expected to display 1D behavior in its ground state. We also present a 1D energy functional, analogous to the Gross-Pitaevskii functional, that correctly describes the energy and density in all the five parameter regions considered here.

E. H. L. is supported by NSF Grant No. PHY-0139984. R. S. is supported by the Austrian Science Fund, FWF and EU network HPRN-CT-2002-0277.

*On leave from Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, A-1090 Vienna, Austria.

- [1] E. H. Lieb and W. Liniger, Phys. Rev. **130**, 1605 (1963).
- [2] E. H. Lieb, Phys. Rev. **130**, 1616 (1963).
- [3] M. D. Girardeau, J. Math. Phys. (N.Y.) **1**, 516 (1960).
- [4] A. Lenard, J. Math. Phys. (N.Y.) **5**, 930 (1964); L. Pitaevskii and S. Stringari, J. Low Temp. Phys. **85**, 377 (1991); M. D. Girardeau, E. M. Wright, and J. M. Trisicari, Phys. Rev. A **63**, 033601 (2001); T. Papenbrock, Phys. Rev. A **67**, 041601(R) (2003).
- [5] A. D. Jackson and G. M. Kavoulakis, Phys. Rev. Lett. **89**, 070403 (2002); S. Komineas and N. Papanicolaou, Phys. Rev. Lett. **89**, 070402 (2002).
- [6] M. Olshanii, Phys. Rev. Lett. **81**, 938 (1998).
- [7] V. Dunjko, V. Lorent, and M. Olshanii, Phys. Rev. Lett. **86**, 5413 (2001); D. S. Petrov, G. V. Shlyapnikov, and J. T. M. Walraven, Phys. Rev. Lett. **85**, 3745 (2000); Ch. Menotti and S. Stringari, Phys. Rev. A **66**, 043610 (2002).
- [8] E. B. Kolomeisky *et al.*, Phys. Rev. Lett. **85**, 1146 (2000); R. K. Bhaduri and D. Sen, Phys. Rev. Lett. **86**, 4708 (2001); E. B. Kolomeisky *et al.*, *ibid.* **86**, 4709 (2001); K. K. Das, Phys. Rev. A **66**, 053612 (2002).
- [9] M. D. Girardeau and E. M. Wright, Phys. Rev. Lett. **87**, 210401 (2001); K. K. Das, M. D. Girardeau, and E. M. Wright, Phys. Rev. Lett. **89**, 110402 (2002).
- [10] K. Bongs *et al.*, Phys. Rev. A **63**, 031602 (2001); A. Görlitz *et al.*, Phys. Rev. Lett. **87**, 130402 (2001); M. Greiner *et al.*, Phys. Rev. Lett. **87**, 160405 (2001); F. Schreck *et al.*, Phys. Rev. Lett. **87**, 080403 (2001).
- [11] D. Blume, Phys. Rev. A **66**, 053613 (2002); G. E. Astrakharchik and S. Giorgini, Phys. Rev. A **66**, 053614 (2002).
- [12] E. H. Lieb, R. Seiringer, and J. Yngvason, math-ph/0305025 [Commun. Math. Phys. (to be published)].
- [13] E. H. Lieb, R. Seiringer, and J. Yngvason, Commun. Math. Phys. **224**, 17 (2001).
- [14] E. H. Lieb, R. Seiringer, and J. Yngvason, Phys. Rev. A **61**, 043602 (2000).
- [15] E. H. Lieb and J. Yngvason, Phys. Rev. Lett. **80**, 2504 (1998).
- [16] E. H. Lieb and R. Seiringer, Phys. Rev. Lett. **88**, 170409 (2002).
- [17] F. J. Dyson, Phys. Rev. **106**, 20 (1957).
- [18] G. Temple, Proc. R. Soc. London A **119**, 276 (1928).

Selecta of Elliott H. Lieb

The Stability of Matter: From Atoms to Stars

Selecta of Elliott H. Lieb

Edited by W. Thirring

With a Preface by F. Dyson

Springer-Verlag Berlin Heidelberg New York 2004

Fourth Edition

Inequalities

Selecta of Elliott H. Lieb

Edited by M. Loss and M. B. Ruskai

Springer-Verlag Berlin Heidelberg New York 2002

Condensed Matter Physics and Exactly Soluble Models

Selecta of Elliott H. Lieb

Edited by B. Nachtergaele, J.P. Solovej and J. Yngvason

Springer-Verlag Berlin Heidelberg New York 2004

Statistical Mechanics

Selecta of Elliott H. Lieb

Edited by B. Nachtergaele, J.P. Solovej and J. Yngvason

Springer-Verlag Berlin Heidelberg New York 2004

Publications of Elliott H. Lieb

1. Second Order Radiative Corrections to the Magnetic Moment of a Bound Electron, *Phil. Mag.* Vol. **46**, 311–316 (1955).
2. A Non-Perturbation Method for Non-Linear Field Theories, *Proc. Roy. Soc.* **241A**, 339–363 (1957).
3. (with K. Yamazaki) Ground State Energy and Effective Mass of the Polaron, *Phys. Rev.* **111**, 728–733 (1958).
4. (with H. Koppe) Mathematical Analysis of a Simple Model Related to the Stripping Reaction, *Phys. Rev.* **116**, 367–371 (1959).
5. Hard Sphere Bose Gas – An Exact Momentum Space Formulation, *Proc. U.S. Nat. Acad. Sci.* **46**, 1000–1002 (1960).
6. Operator Formalism in Statistical Mechanics, *J. Math. Phys.* **2**, 341–343 (1961).
7. (with D.C. Mattis) Exact Wave Functions in Superconductivity, *J. Math. Phys.* **2**, 602–609 (1961).
- ††8. (with T.D. Schultz and D.C. Mattis) Two Soluble Models of an Antiferromagnetic Chain, *Annals of Phys. (N.Y.)* **16**, 407–466 (1961).
- ††\$9. (with D.C. Mattis) Theory of Ferromagnetism and the Ordering of Electronic Energy Levels, *Phys. Rev.* **125**, 164–172 (1962).
- ††\$10. (with D.C. Mattis) Ordering Energy Levels of Interacting Spin Systems, *J. Math. Phys.* **3**, 749–751 (1962).
11. New Method in the Theory of Imperfect Gases and Liquids, *J. Math. Phys.* **4**, 671–678 (1963).
- ††12. (with W. Liniger) Exact Analysis of an Interacting Bose Gas. I. The General Solution and the Ground State, *Phys. Rev.* **130**, 1605–1616 (1963).
- ††13. Exact Analysis of an Interacting Bose Gas. II. The Excitation Spectrum, *Phys. Rev.* **130**, 1616–1624 (1963).
- ††14. Simplified Approach to the Ground State Energy of an Imperfect Bose Gas, *Phys. Rev.* **130**, 2518–2528 (1963).

* means the paper appears in: *Stability of Matter*, ed by W. Thirring, 4th Edition (Springer Berlin Heidelberg 2004)

\$ means the paper appears in: *Inequalities*, ed by M. Loss and M.B. Ruskai (Springer Berlin Heidelberg 2002)

†† means the paper appears in: *Condensed Matter Physics and Exactly Soluble Models*, ed by B. Nachtergaele, J.P. Solovej and J. Yngvason (Springer Berlin Heidelberg 2004)

† means the paper appears in: *Statistical Mechanics*, ed by B. Nachtergaele, J.P. Solovej and J. Yngvason (Springer Berlin Heidelberg 2004)

15. (with A. Sakakura) Simplified Approach to the Ground State Energy of an Imperfect Bose Gas. II. The Charged Bose Gas at High Density, *Phys. Rev.* **133**, A899–A906 (1964).
16. (with W. Liniger) Simplified Approach to the Ground State Energy of an Imperfect Bose Gas. III. Application to the One-Dimensional Model, *Phys. Rev.* **134**, A312–A315 (1964).
- ††17. (with T.D. Schultz and D.C. Mattis) Two-Dimensional Ising Model as a Soluble Problem of Many Fermions, *Rev. Mod. Phys.* **36**, 856–871 (1964).
18. The Bose Fluid, *Lectures in Theoretical Physics, Vol. VIIC*, (Boulder summer school), University of Colorado Press, 175–224 (1965).
- ††19. (with D.C. Mattis) Exact Solution of a Many-Fermion System and Its Associated Boson Field, *J. Math. Phys.* **6**, 304–312 (1965).
- †20. (with S.Y. Larsen, J.E. Kilpatrick and H.F. Jordan) Suppression at High Temperature of Effects Due to Statistics in the Second Virial Coefficient of a Real Gas, *Phys. Rev.* **140**, A129–A130 (1965).
21. (with D.C. Mattis) Book *Mathematical Physics in One Dimension*, Academic Press, New York (1966).
- \$22. Proofs of Some Conjectures on Permanents, *J. of Math. and Mech.* **16**, 127–139 (1966).
23. Quantum Mechanical Extension of the Lebowitz–Penrose Theorem on the van der Waals Theory, *J. Math. Phys.* **7**, 1016–1024 (1966).
24. (with D.C. Mattis) Theory of Paramagnetic Impurities in Semiconductors, *J. Math. Phys.* **7**, 2045–2052 (1966).
25. (with T. Burke and J.L. Lebowitz) Phase Transition in a Model Quantum System: Quantum Corrections to the Location of the Critical Point, *Phys. Rev.* **149**, 118–122 (1966).
26. Some Comments on the One-Dimensional Many-Body Problem, unpublished Proceedings of Eastern Theoretical Physics Conference, New York (1966).
- †27. Calculation of Exchange Second Virial Coefficient of a Hard Sphere Gas by Path Integrals, *J. Math. Phys.* **8**, 43–52 (1967).
- †28. (with Z. Rieder and J.L. Lebowitz) Properties of a Harmonic Crystal in a Stationary Nonequilibrium State, *J. Math. Phys.* **8**, 1073–1078 (1967).
- ††29. Exact Solution of the Problem of the Entropy of Two-Dimensional Ice, *Phys. Rev. Lett.* **18**, 692–694 (1967).
- ††30. Exact Solution of the F Model of an Antiferroelectric, *Phys. Rev. Lett.* **18**, 1046–1048 (1967).
- ††31. Exact Solution of the Two-Dimensional Slater KDP Model of a Ferroelectric, *Phys. Rev. Lett.* **19**, 108–110 (1967).
- ††32. Residual Entropy of Square Ice, *Phys. Rev.* **162**, 162–172 (1967).
33. Ice, Ferro- and Antiferroelectrics, in *Methods and Problems in Theoretical Physics, in honour of R.E. Peierls*, Proceedings of the 1967 Birmingham conference, North-Holland, 21–28 (1970).
34. Exactly Soluble Models, in *Mathematical Methods in Solid State and Superfluid Theory*, Proceedings of the 1967 Scottish Universities’ Summer School of Physics, Oliver and Boyd, Edinburgh 286–306 (1969).

- ††35. Solution of the Dimer Problem by the Transfer Matrix Method, *J. Math. Phys.* **8**, 2339–2341 (1967).
- ††36. (with M. Flicker) Delta-Function Fermi Gas with Two-Spin Deviates, *Phys. Rev.* **161**, 179–188 (1967).
- \$37. Concavity Properties and a Generating Function for Stirling Numbers, *J. Combinatorial Theory* **5**, 203–206 (1968).
- 38. A Theorem on Pfaffians, *J. Combinatorial Theory* **5**, 313–319 (1968).
- ††39. (with F.Y. Wu) Absence of Mott Transition in an Exact Solution of the Short-Range, One-Band Model in One Dimension, *Phys. Rev. Lett.* **20**, 1445–1448 (1968).
- ††40. Two-Dimensional Ferroelectric Models, *J. Phys. Soc. (Japan)* **26** (supplement), 94–95 (1969).
- 41. (with W.A. Beyer) Clusters on a Thin Quadratic Lattice, *Studies in Appl. Math.* **48**, 77–90 (1969).
- 42. (with C.J. Thompson) Phase Transition in Zero Dimensions: A Remark on the Spherical Model, *J. Math. Phys.* **10**, 1403–1406 (1969).
- †*43. (with J.L. Lebowitz) Existence of Thermodynamics for Real Matter with Coulomb Forces, *Phys. Rev. Lett.* **22**, 631–634 (1969).
- 44. Two Dimensional Ice and Ferroelectric Models, in *Lectures in Theoretical Physics, XI D*, (Boulder summer school) Gordon and Breach, 329–354 (1969).
- 45. Survey of the One Dimensional Many Body Problem and Two Dimensional Ferroelectric Models, in *Contemporary Physics: Trieste Symposium 1968*, International Atomic Energy Agency, Vienna, vol. 1, 163–176 (1969).
- 46. Models, in *Phase Transitions*, Proceedings of the 14th Solvay Chemistry Conference, May 1969, Interscience, 45–56 (1971).
- \$47. (with H. Araki) Entropy Inequalities, *Commun. Math. Phys.* **18**, 160–170 (1970).
- ††48. (with O.J. Heilmann) Violation of the Noncrossing Rule: The Hubbard Hamiltonian for Benzene, *Trans. N.Y. Acad. Sci.* **33**, 116–149 (1970). Also in *Annals N.Y. Acad. Sci.* **172**, 583–617 (1971). (Awarded the 1970 Boris Pregel award for research in chemical physics.)
- †49. (with O.J. Heilmann) Monomers and Dimers, *Phys. Rev. Lett.* **24**, 1412–1414 (1970).
- 50. Book Review of “Statistical Mechanics” by David Ruelle, *Bull. Amer. Math. Soc.* **76**, 683–687 (1970).
- 51. (with J.L. Lebowitz) Thermodynamic Limit for Coulomb Systems, in *Systèmes a un Nombre Infini de Degrés de Liberté*, Colloques Internationaux de Centre National de la Recherche Scientifique **181**, 155–162 (1970).
- 52. (with D.B. Abraham, T. Oguchi and T. Yamamoto) On the Anomalous Specific Heat of Sodium Trihydrogen Selenite, *Progr. Theor. Phys. (Kyoto)* **44**, 1114–1115 (1970).
- 53. (with D.B. Abraham) Anomalous Specific Heat of Sodium Trihydrogen Selenite – An Associated Combinatorial Problem, *J. Chem. Phys.* **54**, 1446–1450 (1971).

54. (with O.J. Heilmann, D. Kleitman and S. Sherman) Some Positive Definite Functions on Sets and Their Application to the Ising Model, *Discrete Math.* **1**, 19–27 (1971).
55. (with Th. Niemeijer and G. Vertogen) Models in Statistical Mechanics, in *Statistical Mechanics and Quantum Field Theory*, Proceedings of 1970 Ecole d'Été de Physique Théorique (Les Houches), Gordon and Breach, 281–326 (1971).
- ††56. (with H.N.V. Temperley) Relations between the ‘Percolation’ and ‘Colouring’ Problem and Other Graph-Theoretical Problems Associated with Regular Planar Lattices: Some Exact Results for the ‘Percolation’ Problem, *Proc. Roy. Soc.* **A322**, 251–280 (1971).
57. (with M. de Llano) Some Exact Results in the Hartree-Fock Theory of a Many-Fermion System at High Densities, *Phys. Letts.* **37B**, 47–49 (1971).
- †58. (with J.L. Lebowitz) The Constitution of Matter: Existence of Thermodynamics for Systems Composed of Electrons and Nuclei, *Adv. in Math.* **9**, 316–398 (1972).
59. (with F.Y. Wu) Two Dimensional Ferroelectric Models, in *Phase Transitions and Critical Phenomena*, C. Domb and M. Green eds., vol. 1, Academic Press 331–490 (1972).
- †60. (with D. Ruelle) A Property of Zeros of the Partition Function for Ising Spin Systems, *J. Math. Phys.* **13**, 781–784 (1972).
- †61. (with O.J. Heilmann) Theory of Monomer-Dimer Systems, *Commun. Math. Phys.* **25**, 190–232 (1972). Errata **27**, 166 (1972).
- ††62. (with M.L. Glasser and D.B. Abraham) Analytic Properties of the Free Energy for the “Ice” Models, *J. Math. Phys.* **13**, 887–900 (1972).
- †63. (with D.W. Robinson) The Finite Group Velocity of Quantum Spin Systems, *Commun. Math. Phys.* **28**, 251–257 (1972).
64. (with J.L. Lebowitz) Phase Transition in a Continuum Classical System with Finite Interactions, *Phys. Lett.* **39A**, 98–100 (1972).
65. (with J.L. Lebowitz) Lectures on the Thermodynamic Limit for Coulomb Systems, in *Statistical Mechanics and Mathematical Problems*, Battelle 1971 Recontres, Springer Lecture Notes in Physics **20**, 136–161 (1973).
66. (with J.L. Lebowitz) Lectures on the Thermodynamic Limit for Coulomb Systems, in *Lectures in Theoretical Physics XIV B*, (Boulder summer school), Colorado Associated University Press, 423–460 (1973).
- \$67. Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture, *Adv. in Math.* **11**, 267–288 (1973).
- \$68. (with M.B. Ruskai) A Fundamental Property of Quantum Mechanical Entropy, *Phys. Rev. Lett.* **30**, 434–436 (1973).
- \$69. (with M.B. Ruskai) Proof of the Strong Subadditivity of Quantum-Mechanical Entropy, *J. Math. Phys.* **14**, 1938–1941 (1973).
70. (with K. Hepp) On the Superradiant Phase Transition for Molecules in a Quantized Radiation Field: The Dicke Maser Model, *Annals of Phys. (N.Y.)* **76**, 360–404 (1973).

- ††71. (with K. Hepp) Phase Transitions in Reservoir-Driven Open Systems with Applications to Lasers and Superconductors, *Helv. Phys. Acta* **46**, 573–602 (1973).
- ††72. (with K. Hepp) Equilibrium Statistical Mechanics of Matter Interacting with the Quantized Radiation Field, *Phys. Rev.* **A8**, 2517–2525 (1973).
- ††73. (with K. Hepp) Constructive Macroscopic Quantum Electrodynamics, in *Constructive Quantum Field Theory*, Proceedings of the 1973 Erice Summer School, G. Velo and A. Wightman, eds., Springer Lecture Notes in Physics **25**, 298–316 (1973).
- †\$74. The Classical Limit of Quantum Spin Systems, *Commun. Math. Phys.* **31**, 327–340 (1973).
- 75. (with B. Simon) Thomas-Fermi Theory Revisited, *Phys. Rev. Lett.* **31**, 681–683 (1973).
- \$76. (with M.B. Ruskai) Some Operator Inequalities of the Schwarz Type, *Adv. in Math.* **12**, 269–273 (1974).
- 77. Exactly Soluble Models in Statistical Mechanics, lecture given at the 1973 I.U.P.A.P. van der Waals Centennial Conference on Statistical Mechanics, *Physica* **73**, 226–236 (1974).
- 78. (with B. Simon) On Solutions to the Hartree-Fock Problem for Atoms and Molecules, *J. Chem. Physics* **61**, 735–736 (1974).
- 79. Thomas-Fermi and Hartree-Fock Theory, lecture at 1974 International Congress of Mathematicians, Vancouver. *Proceedings*, **2**, 383–386 (1975).
- \$80. Some Convexity and Subadditivity Properties of Entropy, *Bull. Amer. Math. Soc.* **81**, 1–13 (1975).
- \$81. (with H.J. Brascamp and J.M. Luttinger) A General Rearrangement Inequality for Multiple Integrals, *Jour. Funct. Anal.* **17**, 227–237 (1975).
- \$82. (with H.J. Brascamp) Some Inequalities for Gaussian Measures and the Long-Range Order of the One-Dimensional Plasma, lecture at Conference on Functional Integration, Cumberland Lodge, England. *Functional Integration and its Applications*, A.M. Arthurs ed., Clarendon Press, 1–14 (1975).
- ††83. (with K. Hepp) The Laser: A Reversible Quantum Dynamical System with Irreversible Classical Macroscopic Motion, in *Dynamical Systems*, Battelle 1974 Rencontres, Springer Lecture Notes in Physics **38**, 178–208 (1975). Also appears in *Melting, Localization and Chaos*, Proc. 9th Midwest Solid State Theory Symposium, 1981, R. Kalia and P. Vashishta eds., North-Holland, 153–177 (1982).
- *84. (with P. Hertel and W. Thirring) Lower Bound to the Energy of Complex Atoms, *J. Chem. Phys.* **62**, 3355–3356 (1975).
- *85. (with W. Thirring) Bound for the Kinetic Energy of Fermions which Proves the Stability of Matter, *Phys. Rev. Lett.* **35**, 687–689 (1975). Errata **35**, 1116 (1975).
- †86. (with H.J. Brascamp and J.L. Lebowitz) The Statistical Mechanics of Anharmonic Lattices, in the proceedings of the 40th session of the International Statistics Institute, Warsaw, **9**, 393–404 (1975).

- \$87. (with H.J. Brascamp) Best Constants in Young's Inequality, Its Converse and Its Generalization to More Than Three Functions, *Adv. in Math.* **20**, 151–172 (1976).
- \$88. (with H.J. Brascamp) On Extensions of the Brunn-Minkowski and Prékopa-Leindler Theorems, Including Inequalities for Log Concave Functions and with an Application to the Diffusion Equation, *J. Funct. Anal.* **22**, 366–389 (1976).
- 89. (with J.F. Barnes and H.J. Brascamp) Lower Bounds for the Ground State Energy of the Schroedinger Equation Using the Sharp Form of Young's Inequality, in *Studies in Mathematical Physics*, Lieb, Simon, Wightman eds., Princeton Press, 83–90 (1976).
- \$90. Inequalities for Some Operator and Matrix Functions, *Adv. in Math.* **20**, 174–178 (1976).
- *91. (with H. Narnhofer) The Thermodynamic Limit for Jellium, *J. Stat. Phys.* **12**, 291–310 (1975). Errata *J. Stat. Phys.* **14**, 465 (1976).
- *92. The Stability of Matter, *Rev. Mod. Phys.* **48**, 553–569 (1976).
- 93. Bounds on the Eigenvalues of the Laplace and Schroedinger Operators, *Bull. Amer. Math. Soc.* **82**, 751–753 (1976).
- 94. (with F.J. Dyson and B. Simon) Phase Transitions in the Quantum Heisenberg Model, *Phys. Rev. Lett.* **37**, 120–123 (1976). (See no. 104.)
- \$*95. (with W. Thirring) Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their Relation to Sobolev Inequalities, in *Studies in Mathematical Physics*, E. Lieb, B. Simon, A. Wightman eds., Princeton University Press, 269–303 (1976).
- 96. (with B. Simon and A. Wightman) Book *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*, Princeton University Press (1976).
- 97. (with B. Simon) Thomas-Fermi Theory of Atoms, Molecules and Solids, *Adv. in Math.* **23**, 22–116 (1977).
- †98. (with O.E. Lanford and J.L. Lebowitz) Time Evolution of Infinite Anharmonic Systems, *J. Stat. Phys.* **16**, 453–461 (1977).
- 99. The Stability of Matter, Proceedings of the Conference on the Fiftieth Anniversary of the Schroedinger equation, *Acta Physica Austriaca Suppl.* XVII, 181–207 (1977).
- \$100. Existence and Uniqueness of the Minimizing Solution of Choquard's Non-Linear Equation, *Studies in Appl. Math.* **57**, 93–105 (1977).
- †101. (with J. Fröhlich) Existence of Phase Transitions for Anisotropic Heisenberg Models, *Phys. Rev. Lett.* **38**, 440–442 (1977).
- *102. (with B. Simon) The Hartree-Fock Theory for Coulomb Systems, *Commun. Math. Phys.* **53**, 185–194 (1977).
- 103. (with W. Thirring) A Lower Bound for Level Spacings, *Annals of Phys. (N.Y.)* **103**, 88–96 (1977).
- †104. (with F.J. Dyson and B. Simon) Phase Transitions in Quantum Spin Systems with Isotropic and Non-Isotropic Interactions, *J. Stat. Phys.* **18**, 335–383 (1978).

105. Many Particle Coulomb Systems, lectures given at the 1976 session on statistical mechanics of the International Mathematics Summer Center (C.I.M.E.). In *Statistical Mechanics*, C.I.M.E. 1 Ciclo 1976, G. Gallavotti, ed., Liguore Editore, Naples, 101–166 (1978).
- *106. (with R. Benguria) Many-Body Atomic Potentials in Thomas-Fermi Theory, *Annals of Phys.* (N.Y.) **110**, 34–45 (1978).
- *107. (with R. Benguria) The Positivity of the Pressure in Thomas-Fermi Theory, *Commun. Math. Phys.* **63**, 193–218 (1978). Errata **71**, 94 (1980).
108. (with M. de Llano) Solitons and the Delta Function Fermion Gas in Hartree-Fock Theory, *J. Math. Phys.* **19**, 860–868 (1978).
- †109. (with J. Fröhlich) Phase Transitions in Anisotropic Lattice Spin Systems, *Commun. Math. Phys.* **60**, 233–267 (1978).
- †110. (with J. Fröhlich, R. Israel and B. Simon) Phase Transitions and Reflection Positivity. I. General Theory and Long Range Lattice Models, *Commun. Math. Phys.* **62**, 1–34 (1978). (See no. 124.)
- \$111. (with M. Aizenman and E.B. Davies) Positive Linear Maps Which are Order Bounded on C^* Subalgebras, *Adv. in Math.* **28**, 84–86 (1978).
- \$*112. (with M. Aizenman) On Semi-Classical Bounds for Eigenvalues of Schrödinger Operators, *Phys. Lett.* **66A**, 427–429 (1978).
113. New Proofs of Long Range Order, in *Proceedings of the International Conference on Mathematical Problems in Theoretical Physics* (June 1977), Springer Lecture Notes in Physics, **80**, 59–67 (1978).
- \$114. Proof of an Entropy Conjecture of Wehrl, *Commun. Math. Phys.* **62**, 35–41 (1978).
115. (with B. Simon) Monotonicity of the Electronic Contribution to the Born-Oppenheimer Energy, *J. Phys. B.* **11**, L537–L542 (1978).
- †116. (with O.J. Heilmann) Lattice Models for Liquid Crystals, *J. Stat. Phys.* **20**, 679–693 (1979).
117. (with H. Brezis) Long Range Atomic Potentials in Thomas-Fermi Theory, *Commun. Math. Phys.* **65**, 231–246 (1979).
- *118. The $N^{5/3}$ Law for Bosons, *Phys. Lett.* **70A**, 71–73 (1979).
119. A Lower Bound for Coulomb Energies, *Phys. Lett.* **70A**, 444–446 (1979).
120. Why Matter is Stable, *Kagaku* **49**, 301–307 and 385–388 (1979). (In Japanese).
121. The Number of Bound States of One-Body Schrödinger Operators and the Weyl Problem, *Symposium of the Research Inst. of Math. Sci., Kyoto University*, (1979).
122. Some Open Problems About Coulomb Systems, in *Proceedings of the Lausanne 1979 Conference of the International Association of Mathematical Physics*, Springer Lecture Notes in Physics, **116**, 91–102 (1980).
- \$*123. The Number of Bound States of One-Body Schrödinger Operators and the Weyl Problem, *Proceedings of the Amer. Math. Soc. Symposia in Pure Math.*, **36**, 241–252 (1980).
- †124. (with J. Fröhlich, R.B. Israel and B. Simon) Phase Transitions and Reflection Positivity. II. Lattice Systems with Short-Range and Coulomb Interactions. *J. Stat. Phys.* **22**, 297–347 (1980). (See no. 110.)

125. Why Matter is Stable, Chinese Jour. Phys. **17**, 49–62 (1980). (English version of no. 120).
- †\$126. A Refinement of Simon's Correlation Inequality, Commun. Math. Phys. **77**, 127–135 (1980).
127. (with B. Simon) Pointwise Bounds on Eigenfunctions and Wave Packets in N-Body Quantum Systems. VI. Asymptotics in the Two-Cluster Region, Adv. in Appl. Math. **1**, 324–343 (1980).
128. The Uncertainty Principle, article in *Encyclopedia of Physics*, R. Lerner and G. Trigg eds., Addison Wesley, 1078–1079 (1981).
- \$*129. (with S. Oxford) An Improved Lower Bound on the Indirect Coulomb Energy, Int. J. Quant. Chem. **19**, 427–439 (1981).
- *130. (with R. Benguria and H. Brezis) The Thomas-Fermi-vonWeizsaecker Theory of Atoms and Molecules, Commun. Math. Phys. **79**, 167–180 (1981).
- †131. (with M. Aizenman) The Third Law of Thermodynamics and the Degeneracy of the Ground State for Lattice Systems, J. Stat. Phys. **24**, 279–297 (1981).
- †132. (with J.-R. Bricmont, J. Fontaine, J.L. Lebowitz and T. Spencer) Lattice Systems with a Continuous Symmetry III. Low Temperature Asymptotic Expansion for the Plane Rotator Model, Commun. Math. Phys. **78**, 545–566 (1981).
- †133. (with A.D. Sokal) A General Lee-Yang Theorem for One-Component and Multicomponent Ferromagnets, Commun. Math. Phys. **80**, 153–179 (1981).
- *134. Variational Principle for Many-Fermion Systems, Phys. Rev. Lett. **46**, 457–459 (1981). Errata **47**, 69 (1981).
135. Thomas-Fermi and Related Theories of Atoms and Molecules, in *Rigorous Atomic and Molecular Physics*, G. Velo and A. Wightman, eds., Plenum Press 213–308 (1981).
- *136. Thomas-Fermi and Related Theories of Atoms and Molecules, Rev. Mod. Phys. **53**, 603–641 (1981). Errata **54**, 311 (1982). (Revised version of no. 135.)
137. Statistical Theories of Large Atoms and Molecules, in *Proceedings of the 1981 Oaxtepec conference on Recent Progress in Many-Body Theories*, Springer Lecture Notes in Physics, **142**, 336–343 (1982).
138. Statistical Theories of Large Atoms and Molecules, Comments Atomic and Mol. Phys. **11**, 147–155 (1982).
- *139. Analysis of the Thomas-Fermi-von Weizsäcker Equation for an Infinite Atom without Electron Repulsion, Commun. Math. Phys. **85**, 15–25 (1982).
140. (with D.A. Liberman) Numerical Calculation of the Thomas-Fermi-von Weizsäcker Function for an Infinite Atom without Electron Repulsion, Los Alamos National Laboratory Report, LA-9186-MS (1982).
- *141. Monotonicity of the Molecular Electronic Energy in the Nuclear Coordinates, J. Phys. B.: At. Mol. Phys. **15**, L63-L66 (1982).
142. Comment on “Approach to Equilibrium of a Boltzmann Equation Solution”, Phys. Rev. Lett. **48**, 1057 (1982).
143. Density Functionals for Coulomb Systems, in *Physics as Natural Philosophy: Essays in honor of Laszlo Tisza on his 75th Birthday*, A. Shimony and H. Feshbach eds., M.I.T. Press, 111–149 (1982).

- \$144. An L^p Bound for the Riesz and Bessel Potentials of Orthonormal Functions, *J. Funct. Anal.* **51**, 159–165 (1983).
- \$145. (with H. Brezis) A Relation Between Pointwise Convergence of Functions and Convergence of Functionals, *Proc. Amer. Math. Soc.* **88**, 486–490 (1983).
- *146. (with R. Benguria) A Proof of the Stability of Highly Negative Ions in the Absence of the Pauli Principle, *Phys. Rev. Lett.* **50**, 1771–1774 (1983).
- \$147. Sharp Constants in the Hardy-Littlewood-Sobolev and Related Inequalities, *Annals of Math.* **118**, 349–374 (1983).
- \$148. Density Functionals for Coulomb Systems (a revised version of no. 143), *Int. Jour. Quant. Chem.* **24**, 243–277 (1983). An expanded version appears in *Density Functional Methods in Physics*, R. Dreizler and J. da Providencia eds., *Plenum Nato ASI Series* **123**, 31–80 (1985).
- 149. The Significance of the Schrödinger Equation for Atoms, Molecules and Stars, lecture given at the Schrödinger Symposium, Dublin Institute of Advanced Studies, October 1983, unpublished Proceedings.
- *150. (with I. Daubechies) One Electron Relativistic Molecules with Coulomb Interaction, *Commun. Math. Phys.* **90**, 497–510 (1983).
- 151. (with I. Daubechies) Relativistic Molecules with Coulomb Interaction, in *Differential Equations, Proc. of the Conference held at the University of Alabama in Birmingham, 1983*, I. Knowles and R. Lewis eds., *Math. Studies Series*, **92**, 143–148 North-Holland (1984).
- 152. Some Vector Field Equations, in *Differential Equations, Proc. of the Conference held at the University of Alabama in Birmingham, 1983*, I. Knowles and R. Lewis eds., *Math. Studies Series* **92**, 403–412 North-Holland (1984).
- \$153. On the Lowest Eigenvalue of the Laplacian for the Intersection of Two Domains, *Inventiones Math.* **74**, 441–448 (1983).
- 154. (with J. Chayes and L. Chayes) The Inverse Problem in Classical Statistical Mechanics, *Commun. Math. Phys.* **93**, 57–121 (1984).
- \$155. On Characteristic Exponents in Turbulence, *Commun. Math. Phys.* **92**, 473–480 (1984).
- *156. Atomic and Molecular Negative Ions, *Phys. Rev. Lett.* **52**, 315–317 (1984).
- *157. Bound on the Maximum Negative Ionization of Atoms and Molecules, *Phys. Rev.* **29A**, 3018–3028 (1984).
- 158. (with W. Thirring) Gravitational Collapse in Quantum Mechanics with Relativistic Kinetic Energy, *Annals of Phys. (N.Y.)* **155**, 494–512 (1984).
- 159. (with I.M. Sigal, B. Simon and W. Thirring) Asymptotic Neutrality of Large-Z Ions, *Phys. Rev. Lett.* **52**, 994–996 (1984). (See no. 185.)
- *160. (with R. Benguria) The Most Negative Ion in the Thomas-Fermi-von Weizsäcker Theory of Atoms and Molecules, *J. Phys. B: At. Mol. Phys.* **18**, 1045–1059 (1985).
- \$161. (with H. Brezis) Minimum Action Solutions of Some Vector Field Equations, *Commun. Math. Phys.* **96**, 97–113 (1984).
- \$162. (with H. Brezis) Sobolev Inequalities with Remainder Terms, *J. Funct. Anal.* **62**, 73–86 (1985).

- \$163. Baryon Mass Inequalities in Quark Models, *Phys. Rev. Lett.* **54**, 1987–1990 (1985).
- *164. (with J. Fröhlich and M. Loss) Stability of Coulomb Systems with Magnetic Fields I. The One-Electron Atom, *Commun. Math. Phys.* **104**, 251–270 (1986).
- *165. (with M. Loss) Stability of Coulomb Systems with Magnetic Fields II. The Many-Electron Atom and the One-Electron Molecule, *Commun. Math. Phys.* **104**, 271–282 (1986).
- *166. (with W. Thirring) Universal Nature of van der Waals Forces for Coulomb Systems, *Phys. Rev. A* **34**, 40–46 (1986).
167. Some Ginzburg-Landau Type Vector-Field Equations, in *Nonlinear systems of Partial Differential Equations in Applied Mathematics*, B. Nicolaenko, D. Holm and J. Hyman eds., *Amer. Math. Soc. Lectures in Appl. Math.* **23**, Part 2, 105–107 (1986).
- ††168. (with I. Affleck) A Proof of Part of Haldane’s Conjecture on Spin Chains, *Lett. Math. Phys.* **12**, 57–69 (1986).
- \$169. (with H. Brezis and J-M. Coron) Estimations d’Energie pour des Applications de $\mathbf{R}^3$ a Valeurs dans S^2 , *C.R. Acad. Sci. Paris* **303** Ser. 1, 207–210 (1986).
170. (with H. Brezis and J-M. Coron) Harmonic Maps with Defects, *Commun. Math. Phys.* **107**, 649–705 (1986).
171. Some Fundamental Properties of the Ground States of Atoms and Molecules, in *Fundamental Aspects of Quantum Theory*, V. Gorini and A. Frigerio eds., *Nato ASI Series B*, **144**, 209–214, Plenum Press (1986).
172. (with T. Kennedy) A Model for Crystallization: A Variation on the Hubbard Model, in *Statistical Mechanics and Field Theory: Mathematical Aspects*, *Springer Lecture Notes in Physics* **257**, 1–9 (1986).
173. (with T. Kennedy) An Itinerant Electron Model with Crystalline or Magnetic Long Range Order, *Physica* **138A**, 320–358 (1986).
- ††174. A Model for Crystallization: A Variation on the Hubbard Model, *Physica* **140A**, 240–250 (1986) (Proceedings of IUPAP Statphys 16, Boston).
- ††175. (with T. Kennedy) Proof of the Peierls Instability in One Dimension, *Phys. Rev. Lett.* **59**, 1309–1312 (1987).
- ††176. (with I. Affleck, T. Kennedy and H. Tasaki) Rigorous Results on Valence-Bond Ground States in Antiferromagnets, *Phys. Rev. Lett.* **59**, 799–802 (1987).
- *177. (with H.-T. Yau) The Chandrasekhar Theory of Stellar Collapse as the Limit of Quantum Mechanics, *Commun. Math. Phys.* **112**, 147–174 (1987).
178. (with H.-T. Yau) A Rigorous Examination of the Chandrasekhar Theory of Stellar Collapse, *Astrophys. Jour.* **323**, 140–144 (1987).
- \$179. (with F. Almgren) Singularities of Energy Minimizing Maps from the Ball to the Sphere, *Bull. Amer. Math. Soc.* **17**, 304–306 (1987). (See no. 190.)
180. Bounds on Schrödinger Operators and Generalized Sobolev Type Inequalities, *Proceedings of the International Conference on Inequalities*, University of Birmingham, England, 1987, *Marcel Dekker Lecture Notes in Pure and Appl. Math.*, W.N. Everitt ed., volume 129, pages 123–133 (1991).

- ††181. (with I. Affleck, T. Kennedy and H. Tasaki) Valence Bond Ground States in Isotropic Quantum Antiferromagnets, *Commun. Math. Phys.* **115**, 477–528 (1988).
- 182. (with T. Kennedy and H. Tasaki) A Two Dimensional Isotropic Quantum Antiferromagnet with Unique Disordered Ground State, *J. Stat. Phys.* **53**, 383–416 (1988).
- †183. (with T. Kennedy and B.S. Shastri) Existence of Néel Order in Some Spin 1/2 Heisenberg Antiferromagnets, *J. Stat. Phys.* **53**, 1019–1030 (1988).
- †184. (with T. Kennedy and B.S. Shastri) The XY Model has Long-Range Order for all Spins and all Dimensions Greater than One, *Phys. Rev. Lett.* **61**, 2582–2584 (1988).
- *185. (with I.M. Sigal, B. Simon and W. Thirring) Approximate Neutrality of Large- Z Ions, *Commun. Math. Phys.* **116**, 635–644 (1988). (See no. 159.)
- *186. (with H.-T. Yau) The Stability and Instability of Relativistic Matter, *Commun. Math. Phys.* **118**, 177–213 (1988).
- 187. (with H.-T. Yau) Many-Body Stability Implies a Bound on the Fine Structure Constant, *Phys. Rev. Lett.* **61**, 1695–1697 (1988).
- *188. (with J. Conlon and H.-T. Yau) The $N^{7/5}$ Law for Charged Bosons, *Commun. Math. Phys.* **116**, 417–448 (1988).
- \$189. (with F. Almgren and W. Browder) Co-area, Liquid Crystals, and Minimal Surfaces, in *Partial Differential Equations*, S.S. Chern ed., Springer Lecture Notes in Math. **1306**, 1–22 (1988).
- 190. (with F. Almgren) Singularities of Energy Minimizing Maps from the Ball to the Sphere: Examples, Counterexamples and Bounds, *Ann. of Math.* **128**, 483–530 (1988).
- \$191. (with F. Almgren) Counting Singularities in Liquid Crystals, in *IXth International Congress on Mathematical Physics*, B. Simon, A. Truman, I.M. Davies eds., Hilger, 396–409 (1989). This also appears in: *Symposia Mathematica*, vol. XXX, Ist. Naz. Alta Matem. Francesco Severi Roma, 103–118, Academic Press (1989); *Variational Methods*, H. Berestycki, J.-M. Coron, I. Ekeland eds., Birkhäuser, 17–36 (1990); How many singularities can there be in an energy minimizing map from the ball to the sphere?, in *Ideas and Methods in Mathematical Analysis, Stochastics, and Applications*, S. Albeverio, J.E. Fenstad, H. Holden, T. Lindstrom eds., Cambridge Univ. Press, vol. 1, 394–408 (1992).
- \$192. (with F. Almgren) Symmetric Decreasing Rearrangement can be Discontinuous, *Bull. Amer. Math. Soc.* **20**, 177–180 (1989).
- \$193. (with F. Almgren) Symmetric Decreasing Rearrangement is Sometimes Continuous, *Jour. Amer. Math. Soc.* **2**, 683–773 (1989). A summary of this work (using ‘rectifiable currents’) appears as The (Non)continuity of Symmetric Decreasing Rearrangement in *Symposia Mathematica*, vol. XXX, Ist. Naz. Alta Matem. Francesco Severi Roma, 89–102, Academic Press (1989) and in *Variational Methods*, H. Berestycki, J.-M. Coron, I. Ekeland eds., Birkhäuser, 3–16 (1990).

- ††\$194. Two Theorems on the Hubbard Model, *Phys. Rev. Lett.* **62**, 1201–1204 (1989). Errata **62**, 1927 (1989).
- 195. (with J. Conlon and H.-T. Yau) The Coulomb gas at Low Temperature and Low Density, *Commun. Math. Phys.* **125**, 153–180 (1989).
- \$196. Gaussian Kernels have only Gaussian Maximizers, *Invent. Math.* **102**, 179–208 (1990).
- \$*197. Kinetic Energy Bounds and their Application to the Stability of Matter, in *Schrödinger Operators*, Proceedings Sønderborg Denmark 1988, H. Holden and A. Jensen eds., Springer Lecture Notes in Physics **345**, 371–382 (1989). Expanded version of no. 180.
- *198. The Stability of Matter: From Atoms to Stars, 1989 Gibbs Lecture, *Bull. Amer. Math. Soc.* **22**, 1–49 (1990).
- \$199. Integral Bounds for Radar Ambiguity Functions and Wigner Distributions, *J. Math. Phys.* **31**, 594–599 (1990).
- 200. On the Spectral Radius of the Product of Matrix Exponentials, *Linear Alg. and Appl.* **141**, 271–273 (1990).
- ††\$201. (with M. Aizenman) Magnetic Properties of Some Itinerant-Electron Systems at $T > 0$, *Phys. Rev. Lett.* **65**, 1470–1473 (1990).
- 202. (with H. Siedentop) Convexity and Concavity of Eigenvalue Sums, *J. Stat. Phys.* **63**, 811–816 (1991).
- \$203. (with J.P. Solovej) Quantum Coherent Operators: A Generalization of Coherent States, *Lett. Math. Phys.* **22**, 145–154 (1991).
- 204. The Flux-Phase Problem on Planar Lattices, *Helv. Phys. Acta* **65**, 247–255 (1992). Proceedings of the conference “Physics in Two Dimensions”, Neuchâtel, August 1991.
- 205. Atome in starken Magnetfeldern, *Physikalische Blätter* **48**, 549–552 (1992). Translation by H. Siedentop of the Max-Planck medal lecture (1 April 1992) “Atoms in strong magnetic fields”.
- 206. Absence of Ferromagnetism for One-Dimensional Itinerant Electrons, in *Probabilistic Methods in Mathematical Physics*, Proceedings of the International Workshop Siena, May 1991, F. Guerra, M. Loffredo and C. Marchioro eds., World Scientific pp. 290–294 (1992). A shorter version appears in *Rigorous Results in Quantum Dynamics*, J. Dittrich and P. Exner eds., World Scientific, pp. 243–245 (1991).
- 207. (with J.P. Solovej and J. Yngvason) Heavy Atoms in the Magnetic Field of a Neutron Star, *Phys. Rev. Lett.* **69**, 749–752 (1992).
- *208. (with J.P. Solovej) Atoms in the Magnetic Field of a Neutron Star, in *Differential Equations with Applications to Mathematical Physics*, W.F. Ames, J.V. Herod and E.M. Harrell II eds., Academic Press, pages 221–237 (1993). Also in *Spectral Theory and Scattering Theory and Applications*, K. Yajima, ed., Advanced Studies in Pure Math. **23**, 259–274, Math. Soc. of Japan, Kinokuniya (1994). This is a summary of nos. 215, 216. Earlier summaries also appear in: (a) *Méthodes Semi-Classiques, Colloque international (Nantes 1991)*, Asterisque **210**, 237–246 (1991); (b) *Some New Trends on Fluid Dynamics and Theoretical Physics*, C.C. Lin and N. Hu eds., 149–157, Peking

- University Press (1993); (c) *Proceedings of the International Symposium on Advanced Topics of Quantum Physics, Shanxi*, J.Q. Lang, M.L. Wang, S.N. Qiao and D.C. Su eds., 5–13, Science Press, Beijing (1993).
209. (with M. Loss and R. McCann) Uniform Density Theorem for the Hubbard Model, *J. Math. Phys.* **34**, 891–898 (1993).
210. Remarks on the Skyrme Model, in *Proceedings of the Amer. Math. Soc. Symposia in Pure Math.* **54**, part 2, 379–384 (1993). (Proceedings of Summer Research Institute on Differential Geometry at UCLA, July 8–28, 1990.)
- \$211. (with E. Carlen) Optimal Hypercontractivity for Fermi Fields and Related Noncommutative Integration Inequalities, *Commun. Math. Phys.* **155**, 27–46 (1993).
212. (with E. Carlen) Optimal Two-Uniform Convexity and Fermion Hypercontractivity, in *Quantum and Non-Commutative Analysis*, Proceedings of June, 1992 Kyoto Conference, H. Araki et.al. eds., Kluwer (1993), pp. 93–111. (Condensed version of no. 211.)
- †213. (with M. Loss) Fluxes, Laplacians, and Kasteleyn’s Theorem, *Duke Math. Journal* **71**, 337–363 (1993).
214. (with V. Bach, R. Lewis and H. Siedentop) On the Number of Bound States of a Bosonic N -Particle Coulomb System, *Zeits. f. Math.* **214**, 441–460 (1993).
215. (with J.P. Solovej and J. Yngvason) Asymptotics of Heavy Atoms in High Magnetic Fields: I. Lowest Landau Band Region, *Commun. Pure Appl. Math.* **47**, 513–591 (1994).
216. (with J.P. Solovej and J. Yngvason) Asymptotics of Heavy Atoms in High Magnetic Fields: II. Semiclassical Regions, *Commun. Math. Phys.* **161**, 77–124 (1994).
- *217. (with V. Bach, M. Loss and J.P. Solovej) There are No Unfilled Shells in Unrestricted Hartree-Fock Theory, *Phys. Rev. Lett.* **72**, 2981–2983 (1994).
- \$218. (with K. Ball and E. Carlen) Sharp Uniform Convexity and Smoothness Inequalities for Trace Norms, *Invent. Math.* **115**, 463–482 (1994).
- \$219. Coherent States as a Tool for Obtaining Rigorous Bounds, *Proceedings of the Symposium on Coherent States, past, present and future*, Oak Ridge, D.H. Feng, J. Klauder and M.R. Strayer eds., World Scientific (1994), pages 267–278.
- ††220. The Hubbard model: Some Rigorous Results and Open Problems, in Proceedings of 1993 conference in honor of G.F. Dell’Antonio, *Advances in Dynamical Systems and Quantum Physics*, S. Albeverio et.al. eds., pp. 173–193, World Scientific (1995). A revised version appears in Proceedings of 1993 NATO ASI *The Hubbard Model*, D. Baeriswyl et.al. eds., pp. 1–19, Plenum Press (1995). A further revision appears in *Proceedings of the XIth International Congress of Mathematical Physics*, Paris, 1994, D. Iagolnitzer ed., pp. 392–412, International Press (1995).
221. (with V. Bach and J.P. Solovej) Generalized Hartree-Fock Theory of the Hubbard Model, *J. Stat. Phys.* **76**, 3–90 (1994).

- ††222. Flux Phase of the Half-Filled Band, *Phys. Rev. Lett.* **73**, 2158–2161 (1994).
- \$223. (with M. Loss) Symmetry of the Ginzburg-Landau Minimizer in a Disc, *Math. Res. Lett.* **1**, 701–715 (1994).
- 224. (with J.P. Solovej and J. Yngvason) Quantum Dots, in *Proceedings of the Conference on Partial Differential Equations and Mathematical Physics*, University of Alabama, Birmingham, 1994, I. Knowles, ed., International Press (1995), pages 157–172.
- *225. (with J.P. Solovej and J. Yngvason) Ground States of Large Quantum Dots in Magnetic Fields, *Phys. Rev. B* **51**, 10646–10665 (1995).
- 226. (with J. Freericks) The Ground State of a General Electron-Phonon Hamiltonian is a Spin Singlet, *Phys. Rev. B* **51**, 2812–2821 (1995).
- ††227. (with B. Nachtergaele) Stability of the Peierls Instability for Ring-Shaped Molecules, *Phys. Rev. B* **51**, 4777–4791 (1995).
- 228. (with B. Nachtergaele) Dimerization in Ring-Shaped Molecules: The Stability of the Peierls Instability in *Proceedings of the XIth International Congress of Mathematical Physics*, Paris, 1994, D. Iagolnitzer ed., pp. 423–431, International Press (1995).
- 229. (with B. Nachtergaele) Bond Alternation in Ring-Shaped Molecules: The Stability of the Peierls Instability. In *Proceedings of the conference The Chemical Bond*, Copenhagen 1994, *Int. J. Quant. Chem.* **58**, 699–706 (1996).
- 230. Fluxes and Dimers in the Hubbard Model, in *Proceedings of the International Congress of Mathematicians*, Zürich, 1994, S.D. Chatterji ed., vol. 2, pp. 1279–1280, Birkhäuser (1995).
- *231. (with M. Loss and J. P. Solovej) Stability of Matter in Magnetic Fields, *Phys. Rev. Lett.* **75**, 985–989 (1995).
- *232. (with O. J. Heilmann) Electron Density near the Nucleus of a large Atom, *Phys. Rev A* **52**, 3628–3643 (1995).
- *233. (with A. Iantchenko and H. Siedentop) Proof of a Conjecture about Atomic and Molecular Cores Related to Scott’s Correction, *J. reine u. ang. Math.* **472**, 177–195 (1996).
- ††234. (with L.E. Thomas) Exact Ground State Energy of the Strong-Coupling Polaron, *Commun. Math. Phys.* **183**, 511–519 (1997). Errata **188**, 499–500 (1997).
- \$235. (with L. Cafarelli and D. Jerison) On the Case of Equality in the Brunn-Minkowski Inequality for Capacity, *Adv. in Math.* **117**, 193–207 (1996).
- *236. (with M. Loss and H. Siedentop) Stability of Relativistic Matter via Thomas-Fermi Theory, *Helv. Phys. Acta* **69**, 974–984 (1996).
- ††237. Some of the Early History of Exactly Soluble Models, in *Proceedings of the 1996 Northeastern University conference on Exactly Soluble Models*, *Int. Jour. Mod. Phys. B* **11**, 3–10 (1997).
- *238. (with H. Siedentop and J. P. Solovej) Stability and Instability of Relativistic Electrons in Magnetic Fields, *J. Stat. Phys.* **89**, 37–59 (1997).
- 239. (with H. Siedentop and J-P. Solovej) Stability of Relativistic Matter with Magnetic Fields, *Phys. Rev. Lett.* **79**, 1785–1788 (1997).

240. Stability of Matter in Magnetic Fields, in *Proceedings of the Conference on Unconventional Quantum Liquids, Evora, Portugal, 1996* Zeits. f. Phys. B **933**, 271–274 (1997).
241. Birmingham in the Good Old Days, in *Proceedings of the Conference on Unconventional Quantum Liquids, Evora, Portugal, 1996* Zeits. f. Phys. B **933**, 125–126 (1997).
242. (with M. Loss) *Book Analysis*, American Mathematical Society (1997).
243. Doing Math with Fred, in *In Memoriam Frederick J. Almgren Jr., 1937–1997*, Experimental Math. **6**, 2–3 (1997).
- *244. (with J.P. Solovej and J. Yngvason) Asymptotics of Natural and Artificial Atoms in Strong Magnetic Fields, in *The Stability of Matter: From Atoms to Stars, Selecta of E. H. Lieb*, W. Thirring ed., second edition, Springer Verlag, pp. 145–167 (1997). This is a summary of nos. 207, 208, 215, 216, 224, 225.
245. Stability and Instability of Relativistic Electrons in Classical Electromagnetic Fields, in *Proceedings of Conference on Partial Differential Equations and Mathematical Physics, Georgia Inst. of Tech., March, 1997*, Amer. Math. Soc. Contemporary Math. series, E. Carlen, E. Harrell, M. Loss eds., **217**, 99–108 (1998).
- *246. (with J. Yngvason) Ground State Energy of the Low Density Bose Gas, Phys. Rev. Lett. **80**, 2504–2507 (1998). arXiv math-ph/9712138, mp_arc 97–631.
- †247. (with J. Yngvason) A Guide to Entropy and the Second Law of Thermodynamics, Notices of the Amer. Math. Soc. **45**, 571–581 (1998). arXiv math-ph/9805005, mp_arc 98–339. <http://www.ams.org/notices/199805/lieb.pdf>. See no. 266. This paper received the American Mathematical Society 2002 Levi Conant prize for “the best expository paper published in either the Notices of the AMS or the Bulletin of the AMS in the preceding five years”.
- \$248. (with D. Hundertmark and L.E. Thomas) A Sharp Bound for an Eigenvalue Moment of the One-Dimensional Schroedinger Operator, Adv. Theor. Math. Phys. **2**, 719–731 (1998). arXiv math-ph/9806012, mp_arc 98–753.
- \$249. (with E. Carlen) A Minkowski Type Trace Inequality and Strong Subadditivity of Quantum Entropy, in Amer. Math. Soc. Transl. (2), **189**, 59–69 (1999).
250. (with J. Yngvason) The Physics and Mathematics of the Second Law of Thermodynamics, Physics Reports **310**, 1–96 (1999). arXiv cond-mat/9708200, mp_arc 97–457.
251. Some Problems in Statistical Mechanics that I would like to see Solved, 1998 IUPAP Boltzmann prize lecture, Physica A **263**, 491–499 (1999).
- ††252. (with P. Schupp) Ground State Properties of a Fully Frustrated Quantum Spin System, Phys. Rev. Lett. **83**, 5362–5365 (1999). arXiv math-ph/9908019, mp_arc 99–304.
253. (with P. Schupp) Singlets and Reflection Symmetric Spin Systems, Physica A **279**, 378–385 (2000). arXiv math-ph/9910037, mp_arc 99–404.
- *254. (with R. Seiringer and J. Yngvason) Bosons in a Trap: A Rigorous Derivation of the Gross-Pitaevskii Energy Functional, Phys. Rev A **61**, #043602-1-13 (2000). arXiv math-ph/9908027, mp_arc 99–312.

255. (with J. Yngvason) The Ground State Energy of a Dilute Bose Gas, in *Differential Equations and Mathematical Physics, University of Alabama, Birmingham*, 1999, R. Weikard and G. Weinstein, eds., 295–306, Internat. Press (2000). arXiv math-ph/9910033, mp_arc 99-401.
- *256. (with M. Loss) Self-Energy of Electrons in Non-perturbative QED, in *Differential Equations and Mathematical Physics, University of Alabama, Birmingham*, 1999, R. Weikard and G. Weinstein, eds. 279–293, Amer. Math. Soc./Internat. Press (2000). arXiv math-ph/9908020, mp_arc 99-305.
257. (with R. Seiringer and J. Yngvason) The Ground State Energy and Density of Interacting Bosons in a Trap, in *Quantum Theory and Symmetries, Goslar*, 1999, H.-D. Doebner, V.K. Dobrev, J.-D. Hennig and W. Luecke, eds., pp. 101–110, World Scientific (2000). arXiv math-ph/9911026, mp_arc 99-439.
258. (with J. Yngvason) The Ground State Energy of a Dilute Two-dimensional Bose Gas, *J. Stat. Phys.* **103**, 509–526 (2001). arXiv math-ph/0002014, mp_arc 00-63.
- †259. (with J. Yngvason) A Fresh Look at Entropy and the Second Law of Thermodynamics, *Physics Today* **53**, 32–37 (April 2000). arXiv math-ph/0003028, mp_arc 00-123. See also **53**, 11–14, 106 (October 2000).
260. Lieb-Thirring Inequalities, in *Encyclopaedia of Mathematics, Supplement vol. 2*, pp. 311–313, Kluwer (2000). arXiv math-ph/0003039, mp_arc 00-132.
261. Thomas-Fermi Theory, in *Encyclopaedia of Mathematics, Supplement vol. 2*, pp. 455–457, Kluwer (2000). arXiv math-ph/0003040, mp_arc 00-131.
262. (with H. Siedentop) Renormalization of the Regularized Relativistic Electron-Positron Field, *Commun. Math. Phys.* **213**, 673–684 (2000). arXiv math-ph/0003001 mp_arc 00-98.
263. (with R. Seiringer and J. Yngvason) A Rigorous Derivation of the Gross-Pitaevskii Energy Functional for a Two-dimensional Bose Gas, *Commun. Math. Phys.* **224**, 17–31 (2001). arXiv cond-mat/0005026, mp_arc 00-203.
- *264. (with M. Griesemer and M. Loss) Ground States in Non-relativistic Quantum Electrodynamics, *Invent. Math.* **145**, 557–595 (2001). arXiv math-ph/0007014, mp_arc 00-313.
- *265. (with J.P. Solovej) Ground State Energy of the One-Component Charged Bose Gas, *Commun. Math. Phys.* **217**, 127–163 (2001). Errata **225**, 219–221 (2002). arXiv cond-mat/0007425, mp_arc 00-303.
266. (with J. Yngvason) The Mathematics of the Second Law of Thermodynamics, in *Visions in Mathematics, Towards 2000*, A. Alon, J. Bourgain, A. Connes, M. Gromov and V. Milman, eds., GAFA 2000, no. 1, Birkhauser, p. 334–358 (2000). See no. 247. mp_arc 00-332.
267. The Bose Gas: A Subtle Many-Body Problem, in *Proceedings of the XIII International Congress on Mathematical Physics, London*, A. Fokas, et al. eds. International Press, pp. 91–111, 2001. arXiv math-ph/0009009, mp_arc 00-351.
268. (with J. Freericks and D. Ueltschi) Segregation in the Falicov-Kimball Model, *Commun. Math. Phys.* **227**, 243–279 (2002). arXiv math-ph/0107003, mp_arc 01-243.

269. (with G.K. Pedersen) Convex Multivariable Trace Functions, *Reviews in Math. Phys.* **14**, 1–18 (2002). arXiv math.OA/0107062.
- ††270. (with J. Freericks and D. Ueltschi) Phase Separation due to Quantum Mechanical Correlations, *Phys. Rev. Lett.* **88**, #106401–1–4 (2002). arXiv cond-mat/0110251.
- *271. (with M. Loss) Stability of a Model of Relativistic Quantum Electrodynamics, *Commun. Math. Phys.* **228**, 561–588 (2002). arXiv math-ph/0109002, mp_arc 01-315.
- *272. (with M. Loss) A Bound on Binding Energies and Mass Renormalization in Models of Quantum Electrodynamics, *J. Stat. Phys.* **108**, 1057–1069 (2002). arXiv math-ph/0110027.
- *273. (with R. Seiringer) Proof of Bose-Einstein Condensation for Dilute Trapped Gases, *Phys. Rev. Lett.* **88**, #170409–1–4 (2002). arXiv math-ph/0112032, mp_arc02-115.
274. (with M. Loss) Stability of Matter in Relativistic Quantum Mechanics, in *Mathematical Results in Quantum Mechanics*, Proceedings of QMath8, Taxco, Amer. Math. Soc. Contemporary Mathematics series, pp. 225–238, 2002.
275. (with J. Yngvason) The Mathematical Structure of the Second Law of Thermodynamics, in *Current Developments in Mathematics 2001*, International Press, de Jong et al eds. 2002, pp. 89–129. arXiv math-ph/0204007.
276. (with R. Seiringer, J. P. Solovej and J. Yngvason) The Ground State of the Bose Gas, in *Current Developments in Mathematics 2001*, de Jong et al eds., International Press, 2002, pp. 131–178. arXiv math-ph/0204027, mp_arc 02-183.
277. (with R. Seiringer and J. Yngvason) Poincaré Inequalities in Punctured Domains, *Annals of Math.* **158**, 1067–1080 (2003). arXiv math.FA/0205088.
- *278. (with R. Seiringer and J. Yngvason) Superfluidity in Dilute Trapped Bose Gases, *Phys. Rev. B* **66**, #134529–1–6 (2002). arXiv cond-mat/0205570, mp_arc 02-339.
279. (with F.Y. Wu) The one-dimensional Hubbard model: A reminiscence, *Physica A* **321**, 1–27 (2003). arXiv cond-mat/0207529.
280. (with E. Eisenberg) Polarization of interacting bosons with spin, *Phys. Rev. Lett.* **89**, #220403–1–4 (2002), mp_arc 02-446. arXiv cond-mat/0207042.
281. The Stability of Matter and Quantum Electrodynamics, in Proceedings of the Heisenberg symposium, Munich, Dec. 2001, *Fundamental Physics – Heisenberg and Beyond*, G. Buschhorn and J. Wess, eds., pp. 53–68, Springer (2004). arXiv math-ph/0209034. A modified, up-dated version appears in the *Milan Journal of Mathematics* **71**, 199–217 (2003). A further modification appears in the *Jahresbericht of the German Math. Soc. JB* **106**, 93–110 (2004) (Teubner). arXiv math-ph/0401004.
282. (with R. Seiringer and J. Yngvason) Two-dimensional Gross-Pitaevskii Theory, in: A.G. Litvak (editor). *Progress in Nonlinear Science*, Proceedings of the International Conference Dedicated to the 100th Anniversary of A.A. Andronov, Volume II, pp 582–590, Nizhny Novgorod, Institute of Applied Physics, University of Nizhny Novgorod, 2002.

283. (with R. Seiringer) Equivalent forms of the Bessis-Moussa-Villani Conjecture, *J. Stat. Phys.* **115**, 185–190 (2004). [math-ph/0210027](#).
- *284. (with R. Seiringer and J. Yngvason) One-dimensional Bosons in Three-dimensional Traps, *Phys. Rev. Lett.* **91**, #150401–1–4 (2003). [arXiv cond-mat/0304071](#).
285. (with R. Seiringer and J. Yngvason) One-dimensional Behavior of Dilute, Trapped Bose Gases. *Commun. Math. Phys.* **244**, 347–393 (2004). [arXiv math-ph/0305025](#).
286. (with R. Seiringer) Bose-Einstein Condensation of Dilute Gases in Traps, in *Differential Equations and Mathematical Physics* proceedings of a conference at the Univ. of Birmingham, 2002, Amer. Math. Soc. Contemporary Math. series, Y. Karpeshina, G. Stolz, R. Weikard, Y. Zeng, eds., **327**, 239–250 (2003).
- *287. (with M. Loss) Existence of Atoms and Molecules in Non-Relativistic Quantum Electrodynamics, *Adv. Theor. Math. Phys.* **7**, 667–710 (2003). [arXiv math-ph/0307046](#).
- *288. (with J.P. Solovej) Ground State Energy of the Two-Component Charged Bose Gas. *Commun. Math. Phys.* DOI: 10.1007/s00220-004-1144-1 (in press). [arXiv math-ph/0311010](#), mp_arc 03-490.
289. (with J. Yngvason) The Entropy of Classical Thermodynamics, in *Entropy*, the proceedings of a conference in Dresden, 2000, A. Greven, G. Keller and G. Warnecke eds., Princeton University Press, pp. 147–196 (2003).
290. (with M. Loss) A Note on Polarization Vectors in Quantum Electrodynamics. *Commun. Math. Phys.* (in press) [arXiv math-ph/0401016](#).
- ††291. (with M. Aizenman, R. Seiringer, J.P. Solovej and J. Yngvason) Bose-Einstein Quantum Phase Transition in an Optical Lattice Model. *Phys. Rev. A* **70**, 023612–(1–12) (2004). [arXiv cond-mat/0403240](#).
292. (with E. Carlen) Some Matrix Rearrangement Inequalities, *Annali Matem. Pura Appl.*, (in press). [arXiv math.OA/0402239](#).
- ††293. (with R. Seiringer, J.P. Solovej and J. Yngvason) The Quantum-Mechanical Many-Body Problem: The Bose Gas, in Proceedings of the conference ‘Perspectives in Analysis’, KTH Stockholm, 2003, M. Benedicks, P. Jones and S. Smirnov eds., Springer (in press). [arXiv math-ph/0405004](#).
294. (with M. Loss) The Thermodynamic Limit for Matter Interacting with Coulomb Forces and with the Quantized Electromagnetic Field: I. The Lower Bound. [arXiv math-ph/0408001](#).
295. (with E. Carlen and M. Loss) A Sharp analog of Young’s Inequality on S^N and Related Entropy Inequalities, *Jour. Geom. Anal.* **14**, 487–520 (2004). [arXiv math.FA/0408030](#).