ADVANCES IN QUANTUM CHEMISTRY

Volume 21

Samuel B. Trickey

This is a thematic volume on the Electron Density Functional Method in recognition of the 25th anniversary of the Hohenberg–Kohn–Sham papers.

ADVANCES IN
QUANTUM CHEMISTRY

Volume 21

Density Functional Theory of Many-Fermion Systems

Special Editor

Samuel B. Trickey
Quantum Theory Project, Departments of Physics and Chemistry
University of Florida, Gainesville, Florida

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
San Diego New York Boston London Sydney Tokyo Toronto

This book is printed on acid-free paper. ∞

Academic Press, Inc.
San Diego, California 92101

United Kingdom Edition published by
Academic Press Limited
24-28 Oval Road, London NW1 7DX

Library of Congress Catalog Card Number: 64-8029

ISBN 0-12-034821-7 (alk. paper)

Printed in the United States of America
90 91 92 93 9 8 7 6 5 4 3 2 1

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

R. C. Albers (365), Los Alamos National Laboratory, Los Alamos, New Mexico 87545

J. C. Boettger (365), Los Alamos National Laboratory, Los Alamos, New Mexico 87545

A. M. Boring (365), Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Jerzy Cioslowski (303), Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

B. I. Dunlap (317), Theoretical Chemistry Section, Naval Research Laboratory, Washington, D.C. 20375

D. J. W. Geldart (235), Department of Physics, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada

Roy G. Gordon (341), Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

E. K. U. Gross (255), Department of Physics, University of California, Santa Barbara, Santa Barbara, California 93106

John E. Harriman (27), Department of Chemistry, Theoretical Chemistry Institute, University of Wisconsin, Madison, Madison, Wisconsin 53706

P. C. Hohenberg (7), AT&T Bell Laboratories, Murray Hill, New Jersey 07974

Mark S. Hybertsen (155), AT&T Bell Laboratories, Murray Hill, New Jersey 07974

Jaime Keller (47), División de Ciencias Básicas, Facultad de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 México, D.F., México

Leonard Kleinman (201), Department of Physics, University of Texas, Austin, Texas 78712

W. Kohn (7, 255), Department of Physics, University of California, Santa Barbara, Santa Barbara, California 93106

David C. Langreth (175), Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08855

Richard LeSar (341), Los Alamos National Laboratories, Los Alamos, New Mexico 87545

Mel Levy (69), Department of Chemistry and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118

Steven G. Louie (155), Department of Physics, University of California, Berkeley, Materials and Chemical Sciences Division, Lawrence Berkeley Laboratory, Berkeley, California 94720

Eduardo V. Ludeña (47), Química, Instituto Venezolano de Investigaciones Científicas, Caracas 1020-A, Venezuela

L. N. Oliveira[1] (135), Department of Physics, University of California, Santa Barbara, Santa Barbara, California 93106

John P. Perdew (113), Department of Physics and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118

Mark Rasolt (235), Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

N. Rösch (317), Lehrstuhl für Theoretische Chemie, Technische Universität München, D-8046 Garching, Federal Republic of Germany

Viraht Sahni (201), Department of Physics, Brooklyn College of the City University of New York, Brooklyn, New York 11210

M. Schlter (97), AT&T Bell Laboratories, Murray Hill, New Jersey 07974

L. J. Sham (7, 97), Department of Physics, University of California, San Diego, La Jolla, California 92093

Samuel B. Trickey (1), Quantum Theory Project, Departments of Physics and Chemistry, University of Florida, Gainesville, Florida 32611

G. Vignale (235), Department of Physics, University of Missouri, Columbia, Columbia, Missouri 65211

S. H. Vosko (175), Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada

Weitao Yang (293), Department of Chemistry, University of California, Berkeley, Berkeley, California 94720

[1] Present address: Instituto Fsica Qumica Stildeao Carlos, Universidade de Stildeao Paulo, 13560 Stildeao Carlos, SP-Brazil.

INTRODUCTION

Samuel B. Trickey
Quantum Theory Project
Depts. of Physics and of Chemistry
University of Florida
Gainesville, Florida 32611

From the earliest days of modern quantum mechanics, the notion that the density ought to have a central role in many-Fermion systems has been unrelentingly appealing. The initial manifestation of this siren song was Thomas-Fermi theory/1,2/. As Lieb has shown/3/, TF theory is one of the two limiting cases (the other being the hydrogenic atom) in which simple, exact results about a Fermion system can be obtained. This theme, exact results from density-based theory, is a hallmark of modern density functional theory (DFT) as well. At several points in this volume that theme is pursued.

The early augmentation of TF theory by Dirac/4/ to provide a density-based treatment of exchange and then by von Weizsäcker/5/ to provide a gradient contribution to the kinetic energy, established another theme which continues to this day, namely local approximations and gradient corrections. Again, several authors address the current evolution of that theme in this volume.

The next major step, at least so far as was perceived in the west, was Slater's landmark 1951 paper/5/ which actually introduced the local density approximation (LDA) as we now know it. Contrived as a computationally tractable average of Hartree-Fock (HF) one-electron potentials, it was not obvious for many years (and much controversy) that in fact the Slater approximation is just a simple LDA. The scheme was remarkably successful, particularly given that both its motivation and primary initial use — to predict one-electron energies in solids from the LDA eigenvalues — are now known to have significant formal limitations. That success engendered widespread interest in the theory and also stimulated development of a large body of calculated results which eventually nurtured insight into the formal content of the LDA.

A paper which might have played a role in the development of the theory but did not because of geopolitical considerations was Husimi's 1940 publication of the essential features of what are now called Fermion reduced density matrices/7/. If one could obtain the one- and two-particle reduced density matrices directly, the result would be a many-Fermion theory which is almost as economical procedurally as DFT. For example, all of HF theory for the ground state is determined by the one-particle reduced density matrix, an object which is only somewhat more complicated than its diagonal component, the

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

density. It is therefore unsurprising that Fermion reduced density matrices were re-invented independently in the western literature (in 1954 by Löwdin/8/). It is also no surprise that a portion of that work was devoted to analyzing the connection with Slater's LDA.

Another theme which continues to this day and is addressed in this volume, representability, first emerged from the intense research into the properties of reduced density matrices which began then. The particular form of the problem is N-representability: given a Fermion reduced density matrix, what are the necessary and sufficient conditions that it have a proper N-Fermion wave function (pure or ensemble) as its antecedent? A good picture of the early status of that problem is in Coleman's review from a few years later/9/. Today we are sensitive not only to issues of N-representability but also V-representability [e.g., given a density, what are the necessary and sufficient conditions that it arise from a potential V(**r**) with, perhaps, certain stipulated properties?].

In the light of recent developments in central and eastern Europe, it seems especially sad that Gaspar's 1954 publication of a remarkably prescient version of the DFT total energy in the LDA/10/ went unnoticed in the west for many years. In essence he reversed Slater's procedure. Instead of performing the orbital variation of the Hartree-Fock total energy expression and then making the LDA in the one-electron potential as Slater had, Gaspar made the LDA in the exchange term of the HF total expression, then performed the variation. Although contrived as approximate HF theory, it is obvious retrospectively that Gaspar's scheme has all the structure of modern DFT-LDA, though the paper provided no formal underpinning.

It was this intellectual and historical setting into which Hohenberg, Kohn, and Sham tossed their famous dual bombshell: the Hohenberg-Kohn theorem/11/ and the Kohn-Sham variational procedure/12/. These provide both a formal basis for the focus on the density as the fundamental variational quantity for at least the ground state of many-Fermion systems and a practical procedure for executing the variational calculations. Furthermore, they validated the historical fascination with the density while simultaneously giving a formal setting (and justification, at least for the weakly inhomogeneous system) for the LDA.

This volume is in honor of the twenty-fifth anniversary of that remarkable achievement. The incredible outpouring of papers, involving both applications and formal advances, over those twenty-five years is testimony in itself of the pivotal role those papers have played. So rich is that literature in fact that any effort to provide even a selected bibliography would omit many worthwhile papers while overwhelming the page limits for the volume.

In the interests of brevity, I eschew the ritual summary of contributions to the volume. They can speak more eloquently for themselves than could any paraphrase. From the safety of the editor's chair, let me put forth instead a list of some of the challenges which still confront DFT, in some cases within the LDA, in others not.

1. Although the "third HKS paper"/13/ was rather neglected until recently, it argues clearly that, because the one-electron propagator (Green's function) is a ground state

average it is a functional of the density alone. In spite of both this fact and substantial recent progress (much of it treated in this volume), we do not have a purely DFT formulation for calculating the one-electron propagator and its attendant microscopic dielectric function and mass operator (and this is at the level of the GW approximation/14/, which sets the vertex function to a delta function). The state of the art is to use DFT-LDA quantities as inputs to a direct calculation of the propagator. Furthermore, those calculations are all paramagnetic.

2. As noted at the 1985 Sanibel Symposia/15/, we do not have a systematic understanding of the separated atom limit of the GW procedures that so far have appeared. Since one of the appeals of DFT and LDA is a well-defined separated atom limit, this deficiency is somewhat vexing.

3. LDA separated system limits that are understood are not entirely satisfactory, since the LDA enforces a single Fermi level, hence unphysical orbital occupancies and fractional charges can and do occur at large separations. The result is conceptually (and perhaps practically) disastrous for such problems as atom-surface interactions/16/.

4. There is at least one fractional occupation number problem which seems paradoxical. On the one hand, there is the Janak-Slater theorem/17/ which relates the Kohn-Sham eigenvalues, occupation numbers, and total energy by $\epsilon_j = \partial E/\partial n_j$. This result is customarily used to argue that fractional occupancy will occur only at the Fermi level. However, there is also the generalized Kohn-Sham formulation/18/ which shows that any model one-particle reduced density matrix may occur, hence fractional occupancy may also occur well below the Fermi level. (In particular this would happen if the exact one-particle reduced density matrix for an atom were to be used to construct the explicit kinetic energy term in the Kohn-Sham procedure applied to that atom.)

5. In practical applications of DFT-LDA, there is a well-known but little-understood bond contraction problem/19/. Put roughly, the problem is that the more sophisticated the electron gas model used for the LDA, the more contracted the bond. This is odd in view of the spurious self-repulsion in LDA which, naively at least, might be expected to generate a tendency to bond expansion. Discussions among practitioners sometimes focus on the notion of bond contraction being required in order to overcome self-interaction, but the argument seems to risk anthropomorphism (as if the system somehow "knows" that it must get to a certain binding energy, hence contracts until it achieves it!).

6. Relatively little is known about how to disentangle what Perdew has dubbed the doppelgänger problem, the intertwining of spurious self-interaction and derivative discontinuities (with respect to particle number). In connection with this problem, it is remarkable to note how little is known (other than the work by Talman and co-workers/20/) about the model non-interacting systems which occur in the Kohn-Sham procedure. It is also remarkable how few details are known about what variational trial many-body state implicitly corresponds to a given LDA. For example, it has been clear since at least 1973/21/ that simple "rho-to-the-one-third" LDA does not correspond to the HF trial state for the inhomogeneous system, but to what does it correspond? Reliable knowledge

regarding both the model non-interacting system and the implicit variational state in at least a few well-defined cases might be quite helpful in unmasking the doppelgänger.

7. To my knowledge, no pure DFT treatment of a free negative ion has ever been achieved without artificial constructions like charge-balancing spheres.

In addition to these and similar challenges, there exist such new opportunities as time-dependent DFT. Let us hope that the second quarter-century of DFT will be as fruitful, challenging, and enjoyable as the first has been.

ACKNOWLEDGMENTS

My personal introduction to DFT came as a graduate student in summer 1966, when a classmate, C.C. Rousseau, called to my attention Walter Kohn's lecture in the 1965 Tokyo Summer School book./22/ We were part of a group headed by John Gammel and John Nuttall and I recall that there was much discussion among all of us as to how DFT was related to standard quantum mechanics. Another indelible memory is my personal fascination with the elegant simplicity of the theory. After a season with theory of solid Helium, my career returned to such matters upon becoming a junior (very!) colleague of Per-Olov Löwdin and John C. Slater at Univ. of Florida in January,1969. I am especially grateful therefore to Pierre Hohenberg, Walter Kohn, and Lu Sham for a body of work that has played a major role in most of my career and am also most grateful and honored for having their enthusiastic participation in this project.

Thanks also go to all the authors, several of whom have been extraordinarily patient with the delays in bringing this volume to a close. And, thanks go to my colleagues, current and former, in the Quantum Theory Project. More than once my thinking on DFT, LDA, *et al.* has benefitted from the discipline of their queries and criticisms.

This work was supported in part under Contract DAA L03–87–K-0046 of the U.S. Army Research Office and under U.S. DARPA contract 88–10184.

REFERENCES

1. L.H. Thomas, Proc. Camb. Philos. Soc. 23, 542 (1927)
2. E. Fermi, Rend. Accad. Naz. Lincei 6, 602 (1927)
3. E.H. Lieb, Rev. Mod. Phys. 53, 603 (1981)
4. P.A.M. Dirac, Proc. Camb. Philos. Soc. 26, 376 (1930)
5. C.F. von Weizsäcker, Z.Phys. 96, 431 (1935)
6. J.C. Slater, Phys. Rev. 81, 385 (1951)
7. K. Husimi, Proc. Physico-Math. Soc. Japan 22, 264 (1940)
8. P.O. Löwdin, Phys. Rev. 97, 1474 (1954)
9. A.J. Coleman, Rev. Mod. Phys. 35, 668 (1963)
10. R. Gaspar, Acta Phys. (Hung.) 3, 263 (1954)

11. P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964)

12. W. Kohn and L. Sham, Phys. Rev. 140, A1133 (1965)

13. L. Sham and W. Kohn, Phys. Rev. 145, 561 (1965)

14. L. Hedin and S. Lundqvist, in *Solid State Physics*, F. Seitz, D. Turnbull, and H. Ehrenreich editors, (Academic, NY, 1969),v. 23, p.1.

15. S.B. Trickey, Plenary Lecture, 25th Anniv. Sanibel Symposia, 1985 (unpublished)

16. J.P. Perdew and J.R. Smith, Surf. Sci. Lett. 141, L295 (1984)

17. J.F. Janak Phys. Rev. B 18, 7165 (1978)

18. M. Levy and J.P. Perdew, in *Density Functional Methods in Physics*, R.M. Dreizler and J. da Providencia editors, Plenum, NY, 1985), p.11.

19. H.J.F. Jansen, K.B. Hathaway, and A.J. Freeman, Phys. Rev. B 30, 6177 (1984); J.C. Boettger and S.B. Trickey, Phys. Rev. B 32, 3391 (1985); P. Blaha and K. Schwarz, J. Phys. F.: Met Phys. 17, 899 (1987)

20. For example, K. Ashamar, T.M. Luke, and J.D. Talman, At. Data Nucl. Data Tables 22, 443 (1978)

21. S.B. Trickey, F.R. Green Jr., and F.W. Averill, Phys. Rev. B 8, 4822 (1973)

22. W. Kohn in *Many-Body Theory Part 1*, R. Kubo editor, (Benjamin, NY, 1966), p.73

THE BEGINNINGS
AND SOME THOUGHTS ON THE FUTURE

P.C. Hohenberg

AT&T Bell Laboratories
Murray Hill, NJ 07974

Walter Kohn

Department of Physics
University of California, Santa Barbara
Santa Barbara, CA 93106

L.J. Sham

Department of Physics, B-019
University of California, San Diego
La Jolla, California 92093

ABSTRACT

Reminiscences of how our first work on the density functional theory came about are presented. An attempt is made to describe the motivations and outlook of the authors during these developments. Our account covers

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

the formulation of the two basic theorems expressing the ground state of a many-electron system in terms of the particle density distribution, and of the variational solution in terms of a single-particle Schrödinger equation. In conclusion, some challenges and directions for further progress in this field are indicated.

1. INTRODUCTION

Since, from its beginning, we have had a lot of fun with density functional theory, both doing it and watching other partisans (and even "enemies"), we are quite excited about this 25-year review of its spectacular success story. (We trust that the editor will tactfully suppress failures.) As our own contribution to this volume we offer an account—to the best of our recollection—of how it all began /1,2/; there is no great drama here, but still it is perhaps a case history in which this or that reader may find some interesting or amusing points and from which perhaps she/he may even draw some worthwhile conclusions.

However, we do not want to appear entirely as relics of the past. Therefore, we end the chapter by drawing on our vast experience (beware!) to suggest a few challenging and/or promising directions in which the field might move in the future.

2. HOW IT BEGAN

In the fall of 1963 two of us, Pierre Hohenberg and Walter Kohn, found ourselves inhabiting, on an intermittent basis, the spacious office of Philippe Nozières in the venerable building, 16 rue Lhomond, of the École Normale Supérieure in Paris. Having survived a strenuous term as Department Chairman at the brand new University of California at San Diego, Kohn was there on a Guggenheim-supported sabbatical leave hoping to recuperate. Hohen-

berg had just spent an exciting year at the Institute for Physical Problems in Moscow, after completing his Ph.D. at Harvard in 1962. His training was in many-body theory, dealing primarily with systems with complete translation symmetry, and it was with some trepidation that he accepted Kohn's offer to work with him on the *inhomogeneous* electron gas.

Kohn had been much interested in the electronic screening charge around a charged impurity in metals. His former associates, J.S. Langer and S.H. Vosko, had calculated the charge distribution, $n(r)$, around a weak point charge in a uniform electron gas /3/. With Vosko he had helped to explain measurements in *dilute* copper alloys, $Cu_{1-x}M_x$, $(M = Zn, Al, \ldots)$, of the Cu quadrupole resonance, whose intensity was spectacularly reduced by the screening charges of very small concentrations of multi-valent metals /4/.

In Paris he began to think about the cohesive energies of *concentrated* substitutional metal alloys, $M^{(1)}_{1-x}M^{(2)}_x$ $(x \lesssim \frac{1}{2})$, where the nuclear charge difference $Z^{(2)} - Z^{(1)} \geq 1$. A simple, useful model for such alloys had been the rigid band model: one assumes the same band structure for the alloy as for pure $M^{(1)}$ and (for $Z^{(2)} - Z^{(1)} = 1$) simply "pours" into the available band states an additional electron for each added $M^{(2)}$ atom. This was obviously pretty rough and ready.

A better model for such an alloy was the *average* periodic crystal with nuclear charge $\bar{Z} = (1 - x)Z^{(1)} + xZ^{(2)}$ in each cell, and $\bar{Z}$ electrons/unit cell (the same number as in the rigid band model). In this model the band structure itself depends on x; it obviously gave the right limits for $x = 0$ and 1, and it was easy to see that if the difference, $\Delta \equiv Z^{(2)} - Z^{(1)}$, was treated as a small parameter, this average crystal had the same energy, to first order in Δ, as the actual alloy for all x, no matter what the spatial distribution of $M^{(2)}$ was (including even all many-body effects). The difference in energy between the actual crystal and the average crystal, to first order in Δ, is obtained by first order perturbation theory in $\Delta v(r)$, the difference between

the two external potentials:

$$\begin{aligned}\langle \Delta v(r) \rangle &= \int \Delta v(r) \bar{n}(r) dr \\ &= \sum_{\ell} \left\langle \frac{Z_\ell - \bar{Z}}{|r - R_\ell|} \right\rangle = N(1-x) \left\langle \frac{Z^{(1)} - \bar{Z}}{r} \right\rangle + Nx \left\langle \frac{Z^{(2)} - \bar{Z}}{r} \right\rangle \\ &= N[(1-x)Z^{(1)} + xZ^{(2)} - \bar{Z}] \left\langle \frac{1}{r} \right\rangle = 0, \end{aligned} \tag{1}$$

where ℓ, Z_ℓ, R_ℓ refer to cell ℓ and its nuclear charge and position, and $\langle\rangle$ means expectation value in the average crystal ground state. Of course, as in all first-order calculations, one can legitimately ignore the changes of the ground state wave function and density due to the perturbative potential, $\Delta v(r)$.

From screening of *isolated* impurities, however, it was known that typically most of the screening charge lodged in the central cell. Thus, whereas in the *average* crystal the total charge in each cell is

$$Q \equiv \int_{\text{cell } \ell} \bar{n}(r) dr = \bar{Z}, \textit{all } \ell, \tag{2}$$

in the *real* crystal the charge in cell ℓ with nuclear charge Z_ℓ is

$$Q_\ell \equiv \int_{\text{cell } \ell} n(r) dr \approx Z_\ell \text{ (either } Z^{(1)} \text{ or } Z^{(2)}). \tag{3}$$

Particularly if $Z^{(2)} - Z^{(1)}$ was large (say ≥ 2), there was empirical evidence that the non-uniformity, from cell to cell, of the real density, $n(r)$, would give rise to significant higher-order effects omitted in the lowest-order average periodic crystal.

The importance of the electron density $n(r)$ for the energy became even clearer from the following simple considerations. Let us "turn on" the differences ΔZ_ℓ between Z_ℓ and $\bar{Z}$ by replacing ΔZ_ℓ by $\lambda \Delta Z_\ell$, keeping $\bar{Z}$ fixed, and letting λ grow from 0 to 1. We denote the density distribution corresponding to the value λ by $n_\lambda(r)$. Clearly we have, rigorously,

$$\begin{aligned} E(\lambda = 1) - E(\lambda = 0) &= \int_0^1 \frac{dE}{d\lambda} d\lambda = \int_0^1 d\lambda \int \Delta v(r) n_\lambda(r) dr, \\ &= \int \Delta v(r) \check{n}(r) dr, \end{aligned} \tag{4}$$

where

$$\tilde{n}(r) = \int_0^1 d\lambda n_\lambda(v) dr. \tag{5}$$

To *second* order in ΔZ, the correct energy is obtained from (4) by using

$$\tilde{n}(r) \approx \frac{1}{2}[n_{\lambda=0}(r) + n_{\lambda=1}(r)] = \frac{1}{2}[\bar{n}(r) + n(r)]. \tag{6}$$

At this point the question occurred to Kohn whether a knowledge of $n(r)$ alone determined—at least in principle—the total energy E. Several indications that this might be so came to mind:

For a uniform electron gas, of density n_0, a small sinusoidal perturbing potential, $v_1(q)e^{iq\cdot r}$, produces a small density change, $n_1(q)e^{iq\cdot r}$, proportional to $v_1(q)$,

$$n_1(q) = \alpha(q)v_1(q); \tag{7}$$

here $\alpha(q)$ is the q-dependent density-density response function. For an arbitrary small perturbing potential, $v_1(r)$, the corresponding density change, $n_1(r)$, is obtained by Fourier analysis

$$n_1(r) = \int K_1(r - r')v_1(r')dr', \tag{8}$$

where

$$K_1(r - r') = \frac{1}{(2\pi)^3} \int \alpha(q) e^{iq\cdot(r-r')} dq. \tag{9}$$

This describes the familiar linear screening of a perturbing potential in a metal. In precisely the same way one can obtain the inverse linear relationship.

$$v_1(r) = \int \bar{K}_1(r - r')n_1(r')dr', \tag{10}$$

where

$$\bar{K}_1(r - r') = \frac{1}{(2\pi)^3} \int \frac{1}{\alpha(q)} e^{iq(r-r')} dr'. \tag{11}$$

Thus a small change of density, $n_1(r)$, uniquely determines the external potential, $v_1(r)$, producing it. Formally this inversion can be carried to arbitrary order,

$$\begin{aligned} v(r) &= \int \bar{K}_1(r-r')[n(r')-n_0]dr' \\ &+ \int \bar{K}_2(r-r';r-r'')[n(r')-n_0][n(r'')-n_0]dr'dr'' + \ldots \end{aligned} \quad (12)$$

where $[n(r)-n_0]$ is the change of density caused by the external potential $v(r)$.

Also, the familiar Thomas-Fermi approximation is easily inverted. The conventional self-consistent equations for the density $n(r)$ are

$$n(r) = \gamma[\mu - v_{\text{eff}}(r)]^{3/2} \quad (13)$$

where μ is the chemical potential,

$$\gamma = \frac{1}{3\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2},$$

and

$$v_{\text{eff}}(r) \equiv v(r) + \int \frac{n(r')}{|r-r'|}dr'. \quad (14)$$

They can be trivially solved for $v(r)$ in terms of $n(r)$,

$$v(r) - \mu = -[n(r)/\gamma]^{2/3} - \int \frac{n(r')}{|r-r'|}dr'. \quad (15)$$

Again, $n(r)$ determines $v(r)$ uniquely. (We consider all potentials differing only by an additive constant as equivalent.)

A third example is a single particle in an external potential, $v(r)$. Its wave function satisfies the Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(r) + [v(r) - E]\psi(r) = 0, \quad (16)$$

and

$$n(r) = \psi^2(r); \quad (17)$$

[$\psi(r)$ may be taken as real]. Therefore

$$v(r) - E = \frac{\hbar^2}{2m}\nabla^2\psi(r)/\psi(r) = \frac{\hbar^2}{2m}[\nabla^2 n^{1/2}(r)]/n^{1/2}(r). \tag{18}$$

The combination of these three examples made the following hypothesis plausible: *A given ground state density $n(r)$ is produced by a unique $v(r)$.* Such a simple result (if true), clearly does not depend on the details of the many-electron wave function, and so we looked for a proof based on general principles.

Now, as mentioned above, we were thoroughly aware that the density entered the expression for the *energy* explicitly through the term $\int v(r)n(r)dr$. We therefore turned to the total energy as a quantity which might enable us to prove our conjecture. One of us (W.K.) had used energy variational principles extensively, beginning with his Master's and Doctoral theses. He loved them! So we naturally turned to the question whether the assumption that two different potentials $v'(r)$ and $v(r)$ give rise to the same density, $n(r)$, was consistent with the Rayleigh-Ritz variational principle for the ground state energies E' and E associated with $v'(r)$ and $v(r)$. The Rayleigh-Ritz principle states the following: Let H be the Hamiltonian, consisting of kinetic energy, external potential and interaction energy,

$$H = T + V + U; \tag{19}$$

let Ψ be the ground state wave function associated with $v(r)$; and let Ψ_t be any *different* wave function (apart from an immaterial constant phase factor, $e^{i\theta}$). Then

$$(\Psi_t, H\Psi_t) > (\Psi, H\Psi) = E, \tag{20}$$

where the strong inequality, $>$, rather than $\geq$ enters (at the time we assumed non-degeneracy, whenever required). To highlight the role of the density we wrote (20) in the form

$$(\Psi_t, (T+U)\Psi_t) + \int v(r)n_t(r)dr > (\Psi, (T+U)\Psi) + \int v(r)n(r)dr. \tag{21}$$

In particular, it was tempting to try $\Psi_t = \Psi'$, the ground state associated with $v'(r)$, for which, by assumption, $n_t(r) = n(r)$. Substitution into (10) gave

$$(\Psi', (T+U)\Psi') > (\Psi, (T+U)\Psi). \tag{22}$$

Very strange, since Ψ and Ψ' enter this argument on an equal footing! Reversing their roles gives

$$(\Psi, (T+U)\Psi) > (\Psi', (T+U)\Psi'). \tag{23}$$

These two inequalities are obviously inconsistent. Their addition results in the manifestly false relation

$$0 > 0. \tag{24}$$

Thus the original assumption that the two different potentials $v(r)$ and $v'(r)$ gave rise to the same ground state density $n(r)$ must have been false. *Reductio ad absurdum!* We had proved that a ground state density $n(r)$ uniquely defined the potential $v(r)$ giving rise to it.

On the one hand, especially in view of our examples, this was perhaps not so astonishing. On the other hand, since a knowledge of $v(r)$ completes the definition of the Hamiltonian H, it means that, through H, a knowledge of $n(r)$, a function of three variables, implicitly defines the entire $3N$-dimensional ground state wave function $\Psi(r_1, \cdots r_N)$ and its energy E. For that matter it also defines all excited states $\Psi_j(r_1, \cdots r_N)$ and energies E_j, Green's functions, transition rates,...! It was inebriating!

For a specific example, let $n_1(r)$ and $n_2(r)$ be two periodic densities, one belonging to a metal and the other to an insulator. In principle, the functional forms of these densities alone, coupled with the information that they are ground state densities of electronic systems, allow one to decide which is which—but one needs very subtle spectacles!

Somewhere around this stage of our work, we naturally became very nervous. We began to feel that either we had made some terrible blunder or, if our conclusions were indeed true, they ought to have been discovered decades ago. However, rethinking and restructuring our argument did confirm our result; and in our literature search, which led us to much work on one- and two-particle density matrices (on the whole quite discouraging), and on N-particle phase space theorems, nowhere could we find the simple, direct theorem we had proved.

To Hohenberg, the new results bore interesting analogies with the formal work on Stationary Entropy and Renormalization, which had just been completed by his thesis advisor Paul Martin and Cyrano De Dominicis, also working together in Paris. In two rather forbidding papers /5/ these authors had shown that not only could the one-body potentials be eliminated in favor of the single-particle distribution function, but the program could be formally extended to all vertices, including the anomalous (one- and two-body) potentials which occur in superfluids and superconductors. The present work with Kohn had concentrated on only the first step in this procedure, but the aim was to develop quantitative methods which might lead to useful approximations. Moreover, it took a number of intense but informative discussions with De Dominicis and his colleague Roger Balian to convince them that the procedure worked with the *density* rather than the distribution function.

Since all the information about the ground state was implied by the density $n(r)$, one ought to be able to formulate the Raleigh-Ritz principle entirely in terms of the density, $n(r)$, rather than the many-body wave function, $\Psi(r_1, \cdots r_N)$—another seemingly mad thought. But this turned out to be quite easy. The conventional Rayleigh-Ritz principle said

$$E \leq (\Psi_t, H\Psi_t) = \int v(r) n_t(r) dr + (\Psi_t, (T+U)\Psi_t). \tag{25}$$

We now considered (very naturally in view of the foregoing) those Ψ_t which

were ground states, Ψ', belonging to various external potentials $v'(r)$ with densities $n'(r)$:

$$E \leq (\Psi', H\Psi') = \int v(r)n'(r)dr + (\Psi', (T+U)\Psi'). \tag{26}$$

We had seen that Ψ' was determined by $n'(r)$, or, in other words, was a functional of $n'(r)$. Since T and U were fixed, given operators, this made the entire last term a unique functional of $n'(r)$, which we called $F[n'(r)]$. F was in principle determined for all ground state densities. It could be defined as follows. Let $v'(r)$ have a ground state energy E' and density $n'(r)$, then

$$F[n'(r)] = E' - \int v(r)n'(r)dr. \tag{27}$$

E' and $n'(r)$ can, in principle, be either calculated or measured. This determines F for a particular $n'(r)$. By repeating for all $v'(r)$ one can, in principle, obtain $F[n'(r)]$ for all ground state densities $n'(r)$. At the time, on the basis of our examples, we felt that all "reasonable" functions $n'(r)$ were ground state wave-functions. By "reasonable" we meant obvious characteristics: being positive; $\int n'(r)dr = N'$, an integer; as well as possessing sufficient smoothness. We did not worry about this very much, feeling that exceptions would probably not be physically relevant /6/.

Equation (27) could now be written as follows. Define

$$E_v[n'(r)] \equiv \int v(r)n'(r)dr + F[n'(r)] \tag{28}$$

for all ground state densities, $n'(r)$. Then clearly, for a given $v(r)$,

$$E \leq E_v[n'(r)], \tag{29}$$

and the equality is attained only for the correct $n(r)$ of the ground state corresponding to $v(r)$.

Here was the formal variational principle we had been looking for. For practical purposes we first extracted the mean field Coulomb energy from $F[n]$ and wrote

$$F[n'(r)] = \frac{1}{2}\int \frac{n'(r)n'(r')}{|r-r'|}drdr' + G[n'(r)], \tag{30}$$

which defines G. Various approximations for $G[n']$, gradient expansions, dropping exchange and/or correlation effects, led back to the Thomas-Fermi equation and various refinements. An expansion of n around a constant mean value, n_0, yielded another approximation. Our interest, at that point, was not the actual calculation of $F[n'(r)]$ from a solution of the N-particle Schrödinger equation—a pointless objective since after such a solution the physical problem would be solved by conventional means without further need for anything else. Our interest had rather been to assure ourselves of the *existence* of this functional. We felt—and still feel—that the main significance of our work was that it resulted in a clear and formally rigorous reformulation—now called density functional theory—of quantum mechanical ground states in terms of the density, a vantage point from which existing approximations could be more clearly understood and appraised and from which new and better approximations could be derived.

One incident still remains quite vivid to Hohenberg, and that is the discussion he and Kohn had after they had proved the basic theorems and explored some possible approximation schemes for $F[n]$. The question arose as to what this method might be good for, and Kohn suggested that one could try using it to improve current techniques for calculating band structure in solids. Hohenberg's immediate reaction was to say, "But band structure calculations are horribly complicated, isn't that sort of stuff better left to professionals?" To this (according to Hohenberg) Kohn simply replied, "Young man, I am the Kohn of Kohn and Rostoker!"/7/.

3. IMPROVING THOMAS-FERMI

In the Spring of 1964, Hohenberg was getting ready to go to Bell Laboratories where he worked on other problems, and Kohn returned to La Jolla to find a group of postdocs and visitors eagerly awaiting a first-hand account

of the new work. After his seminar, Kohn suggested to David Mermin that there ought to be a finite temperature version of the Hohenberg-Kohn theorems. Knowing the variational theorem for the grand potential with respect to the correct density matrix, Mermin did the extension in a day /8/. Lu Sham had recently arrived from Cambridge as a post-doc with Kohn. He was given the task of applying the Hohenberg-Kohn theorems to solids, in particular to impurities in semiconductors and metals.

There was at the time in La Jolla a certain amount of skepticism over the theorems. It arose out of the feeling mentioned above, which was to recur frequently especially in quantum chemistry circles, that the one-particle and two-particle density matrices were needed to furnish all the information on the ground state, and that the density, corresponding to the diagonal part of the one-particle density matrix, contained much less information. N. Fukuda, visiting from Tokyo, would construct counter-example after counter-example purporting to demonstrate the non-uniqueness of the potential given a density distribution. The job of resolving these fell to Sham. These exercises did deepen our understanding of the density functional theory. In the end, it was the simplicity of the original *reductio-ad-absurdum* argument which overcame the skepticism for the theorems. The next step remained to find applications for them.

At that point Hohenberg and Kohn had been able to rederive the Thomas-Fermi theory and some known, as well as some new, improvements thereof. We realized, of course, that the density from the Thomas-Fermi theory had some serious defects, particularly the lack of quantum oscillations, such as the shell structure in atoms and the Friedel oscillations due to an impurity. However, the partial summation of the infinite gradient series by HK contained the Kohn singularity and, thus, the physics of the Friedel oscillations. Kohn and Sham /9/ were able, for the one-dimensional non-interacting system, to find a simple way to relate the density to the potential when, in addition

to a slowly varying potential, there was a local perturbation or when the density vanished in some region. The key was the important role which the WKB wave at the Fermi level played in reproducing the density oscillations rather accurately. In principle, the method was easily extended to two or three dimensions by the path integral method. However, finding the classical path for any but a spherical potential was a daunting task. Despite Sham's vaunted skills at solving Cambridge Tripos problems, he found it easier to solve the one-particle Schrödinger equation numerically than to find the classical paths for a billiard ball across a table pock-marked with a periodic array of dents.

4. AND WHY NOT THE HARTREE EQUATIONS?

At the same time we realized, of course, that Hartree and Hartree-Fock equations were much better approximations in producing the density than the Thomas-Fermi method. These also could be derived from a Rayleigh-Ritz principle for the ground state. Furthermore, the Hartree equations could be regarded as a self-consistency problem for the density, $n(r)$, which had emerged as our principal actor. Here are these familiar equations for N electrons.

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v_H(r) - \epsilon_j\right]\psi_j(r) = 0, \tag{31}$$

$$v_H(r) = v(r) + \int \frac{n(r')}{|r - r'|}dr', \tag{32}$$

$$n(r) = \sum_1^N |\psi_j(r)|^2. \tag{33}$$

The self-consistency cycle is as follows: (a) Assume a trial $n(r)$; (b) calculate the Hartree potential $v_H(r)$ from (32); (c) calculate the N lowest single-particle orbitals $\psi_j(r)$ from (31); (d) recalculate $n(r)$ from (33) and check for consistency with the original $n(r)$ of step (a). If not consistent, repeat with a new starting density until self-consistency is obtained.

There was enough family resemblance with the HK theory to make us feel that we ought to be able to extract the Hartree theory, perhaps also Hartree-Fock, and we hoped even for inclusion of correlation effects from the HK variational principle. But how on earth could we extract single-particle equations from the HK variational principle for the total density and energy?

The clue was to consider a gas of *non-interacting* electrons in an external potential, $v(r)$, from the standpoint of the HK theory. Here we had

$$E_v[n(r)] = \int v(r)n(r)dr + F_s[n(r)], \tag{34}$$

where F_s was the F-functional for non-interacting (or single-particle) electrons. Of course we knew, independently, that for these non-interacting electrons the density, $n(r)$, could be obtained by solving the single-particle wave equation,

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v(r) - \epsilon_j\right]\psi_j(r) = 0, \tag{35}$$

and calculating

$$n(r) \equiv \sum_1^N |\psi_j(r)|^2. \tag{36}$$

For this non-interacting system

$$F_s[n(r)] = \sum_1^N \epsilon_j - \int v(r)n(r)dr = T_s[n(r)], \tag{37}$$

where, evidently,

$$T_s[n(r)] \equiv \text{kinetic energy of a non-interacting electron gas of density } n(r), \tag{38}$$

in the appropriate external potential $v(r)$. The HK analysis, applied to this *non-interacting* system demonstrated that $T_s[n(r)]$ was indeed a universal functional of $n(r)$ only; for $n(r) \to v(r) \to \psi_j(r) \to T_s = \frac{\hbar^2}{2m}\sum_j |\nabla\psi_j|^2]$. So now we saw explicitly that the variational principle

$$E_v[n'(r)] \equiv \int v(r)n'(r)dr + T_s[n'(r)], \tag{39}$$

$$\delta E_v[n'(r)] \equiv \left[v(r) + \frac{\delta T_s[n'(r)]}{\delta n'(r)}\right]\delta n'(r) = 0 \tag{40}$$

was solved by the equations (35), (36). We were edging up!

Going back to the *interacting* system, from which we wanted to extract the Hartree equations (or better) it was natural to write $E_v[n'(r)]$ in a form similar to (39). This meant extracting $T_s[n'(r)]$, the kinetic energy for *non-interacting* electrons from $F[n'(r)]$ for *interacting* electrons [Note: *not* $T[n(r)]$, the kinetic energy, $\langle \sum \frac{\hbar^2}{2m}\nabla_j^2 \rangle$, for *interacting* electrons of density $n(r)$]. So we wrote

$$F[n'(r)] \equiv T_s[n'(r)] + \frac{1}{2}\int \frac{n'(r)n'(r')}{|r-r'|}drdr' + E_{xc}[n'(r)]. \tag{41}$$

Since, by the HK theory, both $F[n'(r)]$ and $T_s[n'(r)]$ are functionals of the density which are, in principle, known, Eq. (41) *defines* the last term, E_{xc}, which we called exchange-correlation functional.

The HK variational principle gave

$$\delta E_v[n'(r)] = \delta\left\{\int n'(r)v(r)dr + T_s[n'(r)] + \frac{1}{2}\int \frac{n'(r)n'(r')}{|r-r'|}dr' + E_{xc}[n'(r)]\right\}$$

$$= \left[v(r) + \int \frac{n'(r')}{|r-r'|}dr' + \frac{\delta E_{xc}}{\delta n'(r)} + \frac{\delta T_s}{\delta n'(r)}\right]\delta n'(r) = 0. \tag{42}$$

This has the same structure as Eq. (40) with the replacement

$$v(r) \to v_{\text{eff}}(r) \equiv v(r) + \int \frac{n'(r')}{|r-r'|}dr' + \frac{\delta E_{xc}}{\delta n'(r)}. \tag{43}$$

Since Eq. (40) was solved by Eqs. (35) and (36), Eq. (42) could be solved by

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v_{\text{eff}}(r) - \epsilon_j\right]\psi_j = 0, \tag{44}$$

$$n'(r) = \sum_1^N |\psi_j(r)|^2, \tag{45}$$

$$v_{\text{eff}}(r) = v(r) + \int \frac{n'(r')}{|r-r'|}dr' + v_{xc}(r), \tag{46}$$

$$v_{xc}(r) \equiv \frac{\delta E_{xc}}{\delta n'(r)}. \tag{47}$$

These self-consistent equations are very analogous to the Hartree equations and, like the latter, must be solved self-consistently: One starts with a trial-density $n(r)$. Assuming that the functional $E_{xc}[n(r)]$ is known (see below),

one constructs v_{xc} from (47), then v_{eff} from (46), solves (44) and recalculates $n(r)$ from (45). This cycle is repeated until self-consistency is achieved. Evidently, the Hartree equations are obtained by ignoring v_{xc} altogether, in Eq. (46). However, much more interesting, if v_{xc} is calculated from the *correct* E_{xc}, the system of Hartree-like equations, (44)-(47), yields—as we were rather amazed to note—the *exact* ground state density $n(r)$ and—by Eqs. (28) and (41)—the *exact* many-body ground state energy,

$$\begin{aligned} E &= T_s[n] + \int v(r)n(r)dr + \frac{1}{2}\int \frac{n(r)n(r')}{|r-r'|}drdr' + E_{xc}[n(r)], \\ &= \sum_1^N \epsilon_j - \frac{1}{2}\int \frac{n(r)n(r')}{|r-r'|} + E_{xc}[n(r)] - \int v_{xc}(r)n(r)dr. \end{aligned} \tag{48}$$

At the time we simply assumed that for all "reasonable" systems the needed quantities, $T_s[n], E_{xc}[n]$ and their first variations existed. Much later, Levy and Lieb /6/ showed interesting examples, for ground states with degeneracies, of reasonable-looking densities for which these quantities did *not* exist.

It was clear that there was no theoretical underpinning for taking the energy eigenvalues, ϵ_j, from the Kohn-Sham equation (44) to be the true excited state energies. We were, however, able to construct a local density approximation for the self-energy operator to obtain the quasi-particle energies of solids /11/. Nevertheless, the temptation of using the simpler (44) for the band structures in solids proved irresistible. This has yielded useful first approximations for most energy band calculations. Only recently, after it was clearly established that the ϵ_j gave inaccurate band gaps in semiconductors and insulators, have serious efforts begun to calculate bands from Green's functions for which DFT provides a starting point.

Extensions to systems with external potentials $v(r)$ *and* magnetic fields, $H_z(r)$, where $n(r)$ was replaced by $n_\uparrow(r)$ and $n_\downarrow(r)$ were obvious, and we made them. We included only the effects of spin coupling at the time. It was also straightforward to obtain formally exact self-consistent single particle equations for finite temperature ensembles.

5. PRACTICAL APPROXIMATIONS

Up to this point the theory had been purely formal. For useful applications we had to find sufficiently simple and yet effective approximations for $E_{xc}[n(r)]$. Of course, knowing that even setting $E_{xc}[n(r)] = 0$ leads to the Hartree equations, which had a long history of yielding valuable results, we felt that even rather crude approximations to $E_{xc}[n(r)]$ would yield a practically useful theory. HK had already used rather obvious gradient expansions for $G[n]$. Now we naturally used them for $E_{xc}[n]$:

$$E_{xc}[n(r)] = \int \epsilon_{xc}(n(r))n(r)dr + O(|\nabla n(r)|^2), \tag{49}$$

where $\epsilon_{xc}(n)$ is the exchange-correlation energy per particle of a *uniform* gas of density $n(r)$. Using the approximation in (49), we obtained the exchange-correlation potential to be the exchange-correlation part of the chemical potential of the homogeneous electron gas, $\mu_{xc}(n)$.

We recognized that the exchange part of this potential, $\mu_x(n)$, was two-thirds of the famous Slater local exchange potential /10/. Slater derived his approximation in two steps: (1) the wave functions in the exchange operator were replaced by plane waves, and (2) the exchange was averaged over the entire Fermi sea. Our result corresponded to taking the exchange at the Fermi level in step (2). We were reassured by the thermodynamic argument that the density was determined by the chemical potential in a slowly varying external potential.

Keeping only the leading term in (49) gave the so-called local density approximation (LDA) on which at least 90% of all applications to date have been based. Gradient expansions for systems of slowly varying density and perturbation expansions for systems of nearly uniform density were obvious extensions—which, in practice, turned out not to be very useful.

The first chapters of this story had been written.

6. LOOKING AHEAD

In the intervening years DFT has mushroomed, both in important further theoretical developments and in applications. Much of this story is told in the other chapters of this volume.

What of the future? In recent years we have increasingly seen DFT used fruitfully as a starting point or in conjunction with other methods, such as Green's function theory, hypernetted chain equations and Monte Carlo or molecular dynamics methods. Perhaps we will see more of these "marriages" in the future.

Time-dependent density functional theory has been formally developed, but it has not yet found a practical incarnation of an ease and utility comparable to LDA for stationary ground and excited states and thermal ensembles.

The practical implementations of DFT all take the uniform electron gas as their starting point for approximate representations of exchange-correlation effects. But in bounded systems (atoms, molecules, surfaces, etc.) the boundary region(s), where the density tends to zero and where the Kohn-Sham wave-functions *evanesce* rather than oscillate, are *essentially* different from a region of a uniform electron-gas, even if the density does not vary rapidly. We hope that good theories for these tails and for the transition regions linking them to "good" interior regions will soon be devised. Another potential area for future progress is the problem of strong local electron correlation, occurring most notably in the recently discovered high temperature superconducting oxides. The Hubbard model has been extensively used to study the effects of strong intrasite correlation. Density functionals can be used to derive this model and corrections to it, starting from a more realistic Hamiltonian.

Of course, a number of "classic" DFT developments are still fairly recent (excited states, diamagnetism, superconductivity) and we expect consider-

able future applications and experience—applied DFT is always partly an art!

Interestingly, there is an analogy in the formal aspects of our original work /1/ with modern developments in the description of spatially structured states in systems far from thermal equilibrium /12/. The weakly inhomogeneous approximation, valid for $|n(r) - n_0| \ll n_0$, corresponds to the Landau expansion of unstable systems near threshold, and leads to the well-known "amplitude equations." The slowly varying approximation, $|\nabla n|/n \ll n^{1/3}$, corresponds to the gradient expansions which are carried out in nonequilibrium systems far from threshold, and lead to "phase equations."

Finally, stepping back for a wider look, we expect future uses of DFT concepts for other physical problems involving *spatially inhomogeneous environments.* Examples might be the Landau theory of Fermi liquids for inhomogeneous systems; phase transitions in systems with surfaces; Landau-Ginzburg theory of superconductivity; and localized or generalized spins in inhomogeneous magnetic fields.

ACKNOWLEDGEMENTS

W.K. is supported in part by NSF Grant No. DMR 87-03434 and L.J.S. is supported in part by NSF Grant No. DMR 88-15068.

REFERENCES

1. Hohenberg, P. and Kohn, W. (1964). Phys. Rev. **136**, B684.
2. Kohn, W. and Sham, L.J. (1965). Phys. Rev. **140**, A1133.
3. Langer, J.S. and Vosko, S.H. (1959). J. Phys. Chem. Solids **12**, 196.
4. Kohn, W. and Vosko, S.H. (1960). Phys. Rev. **119**, 912.
5. De Dominicis, C. and Martin, P.C. (1964). J. Math. Phys. **5**, 14 and 31.
6. In the meantime, M. Levy [(1982). Phys. Rev. A **26**, 1200] and E. Lieb [(1983). Int. J. Quan. Chem. **24**, 243] have explicitly exhibited a very small class of "reasonable" $n'(r)$ which are not ground state

densities for *any* $v(r)$, or, in current terminology, not v-representable. In practice, these exceptions have not proved to be any obstacle to density functional calculations for physical systems. See also M. Levy's contribution to this volume.
7. Kohn, W. and Rostoker, N. (1954). Phys. Rev. **94**, 1111.
8. Mermin, N.D. (1965). Phys. Rev. **137**, A1441.
9. Kohn, W. and Sham. L.J. (1965). Phys. Rev. **137**, A1697.
10. Slater, J.C. (1951). Phys. Rev. **81**, 385.
11. Sham, L.J. and Kohn, W. (1966). Phys. Rev. **145**, 561.
12. Cross, M.C. and Hohenberg, P.C. (1987) In *Fluctuations and Stochastic Phenomena in Condensed Matter*, Lecture Notes in Physics 268, L. Garrido, Ed., Springer Verlag, New York; and Rev. Mod. Phys. (to be published).

DENSITY AND DENSITY MATRICES IN DENSITY FUNCTIONAL THEORY

John E. Harriman

Department of Chemistry and Theoretical Chemistry Institute
University of Wisconsin-Madison
Madison, Wisconsin 53706

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

1. INTRODUCTION

In its simplest form, the Hohenberg-Kohn theorem /1/ establishes a one-to-one relationship between an "external" potential $v(\mathbf{r})$ and a ground state density $\rho(\mathbf{r})$. It is thus also possible to say that v is a *functional* of ρ: in principle, given ρ one could determine v, and then the ground state energy E_0. It follows that the E_0 is also a functional of ρ:

$$E_0 = E[\rho] = F[\rho] + \int v(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r} \tag{1}$$

where $F[\rho]$ is "universal" in the sense that it does not depend on v. Hohenberg and Kohn went on to show that

$$\tilde{E} = E[\tilde{\rho}] = \int v(\mathbf{r})\tilde{\rho}(\mathbf{r})\, d\mathbf{r} + F[\tilde{\rho}] \geq E_0 = E[\rho] \tag{2}$$

where $\tilde{\rho}$ is any proper trial density. One could thus approximate E and ρ by finding the $\tilde{\rho}$ that minimizes $\tilde{E}$. This is the basis for density functional theory (DFT). Note that the density is assumed to have a fixed normalization. We assume ρ integrates to n, the number of electrons in the system. An alternative normalization is sometimes convenient and we will use $d = \rho/n$, which integrates to 1.

Two questions immediately arose: what, explicitly, is the functional $F[\rho]$ and how is the set of acceptable trial densities $\{\tilde{\rho}\}$ characterized? An acceptable density must be n representable /2/, i.e. it must be a possible density for some n-electron system, and it must be v representable /3/, i.e. there must be some local potential $v(\mathbf{r})$ for which it is the ground state density. Most of conventional DFT has addressed the first question by using an orbital approach for the kinetic energy and approximations based on electron gas theory for the exchange-correlation energy. The second question has been largely ignored.

As will be shown in the next section, any reasonable density is n representable. The v-representability question is much more difficult, but fortunately it can be effectively avoided for purposes of formal discussion. This is accomplished by the Levy-Lieb constrained search functional /4, 5/,

$$F_L[\rho] = \min_{\Psi \to \rho} \int \Psi^*(\hat{T} + \hat{U})\Psi\, d\tau\ , \tag{3}$$

where $\hat{T}$ and $\hat{U}$ are the kinetic energy and electron-electron interaction terms, respectively, in the Hamiltonian and the set $\{\Psi|\Psi \rightarrow \rho\}$ consists of all normalized n-electron wavefunctions that give the density ρ. A number of mathematical technicalities, such as whether the minimum should be an infimum and some subtle conditions on ρ, will be ignored here.

While the $F_L[\rho]$ is well-defined and provides an algorithm in terms of which a calculation could in principle be done, there was no explicit characterization of the set of wavefunctions consistent with a given density. Such a characterization becomes possible if the functional is extended as suggested by Valone /6/: instead of considering only wavefunctions, one considers all n-electron density matrices consistent with the given density. The characterization of the set of one-electron reduced density matrices that are ensemble n representable, i.e. can be obtained by integration over the coordinates of $n-1$ electrons in an n-electron density matrix, has been known for some time /2/. It is also straightforward, as will be shown below, to characterize the set of one-electron reduced density matrices (or one-matrices) that correspond to a given density. While the constrained search approach is of prime importance for formal investigation, it does not lead to efficient actual calculations. It is much easier to do a single, direct variational calculation than to do a constrained search at each step in a variation based on the density.

In addition to the computational advantages that may be associated with the use of a density rather than a wavefunction, density functional theory provides the attractive feature that the density is an experimental observable while a wavefunction is not. The momentum density is also observable and provides additional information about the system. The possibility of a theory based on both coordinate and momentum densities will be considered.

2. BASIS SETS AND MODELS

Much of conventional density functional theory has involved numerical representations of the density. In some cases, particularly for molecular systems, this approach has been found unsatisfactory and basis sets similar to those used

in other molecular electronic structure methods have been introduced. Perhaps more important for present purposes, the use of a basis set avoids many of the formal mathematical difficulties that we do not expect to be important in actual calculations anyway. This is particularly true in the case of finite basis sets.

The introduction of a finite basis set defines a model in which operators are replaced by finite-dimensional matrices. Model problems can be solved exactly, to within the numerical precision of the computer being used, and it is commonly assumed that by increasing the size of the basis set and judiciously choosing the basis functions one can build a model problem that approximates as nearly as desired the physical problem of interest.

The introduction of an explicit basis set suggests the introduction of creation and annihilation operators and the use of a second-quantized formalism. This formalism deals with some aspects of the problem with such efficiency that other aspects may be overlooked. For present purposes we will want to call attention to and deal explicitly with certain details that are usually included only implicitly in a second quantized treatment. We will thus not use such a treatment here.

2.1. Particle Number and Order

Consider the Hamiltonian operator, in the ordinary coordinate representation of the Schrödinger picture, for an n-electron system. It can be written as $\hat{H} = \hat{T} + \hat{V} + \hat{U}$ where, in atomic units,

$$\hat{T} = \sum_{j=1}^{n} \left(-\frac{1}{2} \nabla_j^2 \right) , \qquad \hat{V} = \sum_{j=1}^{n} v(\mathbf{r}_j) , \qquad \hat{U} = \sum_{k>j=1}^{n} \frac{1}{r_{jk}} \tag{4}$$

are the kinetic energy, external potential, and electron-electron interaction contributions, respectively. Each of these is an n-electron operator in the sense that it is an operator acting on n-electron functions, but it is also common to refer to $\hat{T}$ and $\hat{V}$ as one-electron operators and to $\hat{U}$ as a two-electron operator, because they involve the coordinates of only one or two electrons at a time. To resolve this ambiguity a more precise terminology will be used here: $\hat{T}$ and $\hat{V}$ will be referred to as being of *order* 1 while $\hat{U}$ is of *order* 2. All will be called n-electron operators.

2.2. Fundamental Hilbert Spaces

Assume that the model problem is defined in terms of an orbital basis set $\{\chi_j,\ j = 1,\ldots,b\}$. Orthonormality will be assumed for convenience, but the only essential requirement is linear independence. These functions span, and thus define, a Hilbert space ${\mathcal{H}_1}^0$. (In the notation used here for spaces, a numerical subscript is a particle number. The superscript 0 in this case refers to the fact that this is an orbital rather than a spinorbital space.) From the b orbitals we can construct $2b$ spinorbitals that span a spinorbital space $\mathcal{H}_1$.

For any p between 2 and $2b$ we can define a determinantal function $\Phi_{K_p} = \hat{A}_p \chi_{k_1} \cdots \chi_{k_p}$ where the index set $K_p = k_1,\ldots,k_p$ labels a set of occupied orbitals and $\hat{A}_p$ is the p electron, normalizing antisymmetrizer. There are $\binom{2b}{p}$ such functions possible and we thus have a sequence of Hilbert spaces $\mathcal{H}_p$ of dimension $\binom{2b}{p}$.

2.3. Matrices and Restricted Operators

2.3.1. Matrices

In a configuration interaction (CI) calculation we would (at least in principle) construct the full CI Hamiltonian matrix $\mathbf{H}$ with elements

$$H_{K_n L_n} = \int \Phi^*_{K_n} \hat{H} \Phi_{L_n}\, d\tau. \tag{5}$$

All the information required to construct this $\binom{2b}{n} \times \binom{2b}{n}$ matrix is contained in the much smaller matrices of the one and two electron integrals

$$\begin{aligned} f_{jk} &= \int \chi_j^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 + v(\mathbf{r}) \right) \chi_k(\mathbf{r})\, d\mathbf{r} \\ g_{jklm} &= \iint \chi_j^*(\mathbf{r}_1) \chi_k(\mathbf{r}_2) \frac{1}{r_{12}} \chi_l(\mathbf{r}_1) \chi_m(\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2\ . \end{aligned} \tag{6}$$

If spin and symmetry-adapted configuration functions are used, rather than single determinants, the construction of $\mathbf{H}$ from $\mathbf{f}$ and $\mathbf{g}$ or of a reduced density matrix from the CI wavefunction expansion coefficients is more complicated. It is the triumph of modern CI methods that this construction is still possible in highly efficient ways.

2.3.2. Restriction of Operators

One can restrict attention to the model problem defined by a particular basis while still dealing, at least conceptually, with operators rather than matrices. For any general n-electron operator $\hat{W}$, the model-problem operator $\hat{W}|_{\mathcal{H}_n}$ is an integral operator with kernel

$$W(N;N') = \sum_{J_n,K_n} W_{J_nK_n} \Phi_{J_n}(N) \Phi^*_{K_n}(N') \tag{7}$$

where N stands for the set of coordinates of n electrons and N' represents a similar set with a prime on each coordinate. This is equivalent to defining the model operator by restricting the domain of $\hat{W}$ to $\mathcal{H}_n$ and projecting its range onto $\mathcal{H}_n$, i.e. $\hat{W}|_{\mathcal{H}_n} = \hat{P}[\mathcal{H}_n]\hat{W}\hat{P}[\mathcal{H}_n]$. The integral kernel formulation has the advantage of making clear the relationship between density matrices, normally expressed in terms of integral kernels, and other operators.

The reduction superoperator /7/ defined initially for density matrices can be applied to any operator. For any p and q with $1 \leq q < p \leq n$, an operator kernel $D^{(p)}(P;P')$ is reduced to an operator with kernel $D^{(q)}(Q;Q')$ by dividing the coordinate set P into subsets (Q,R) and integrating over R:

$$D^{(q)}(Q;Q') = \hat{L}^q_p D^{(p)}(P;P') = C(pqn) \int D^{(q)}(Q,R;Q',R)\, dR\,. \tag{8}$$

The coefficient $C(pqn)$ depends on the normalization convention in effect for reduced density matrices.

2.4. Local and Nonlocal Operators

A distinction is made between local density functionals and nonlocal functionals. In general, the concept of locality is representation dependent: the kinetic energy operator is nonlocal in the coordinate representation but would be local in the momentum representation. More often in DFT, one is concerned with nonlocality in the sense of an integral operator like the self-consistent field (SCF) exchange operator and its possible approximation by a local operator. The concept of locality must be reexamined in a basis set model.

In lattice models or related treatments, the analog of a local operator is taken to be a diagonal matrix. This is not an appropriate identification for a basis set

model of the type being considered here. We certainly think of $v(\mathbf{r})$ as a local quantity, but in general the matrix corresponding to this potential function will not be diagonal. On the other hand, it is always possible to choose a basis so that the matrix of *any* operator, such as the kinetic energy, will be diagonal. A more useful definition of locality in a matrix model is thus that $\mathbf{f}$ is local if there is a function $f(\mathbf{r})$ such that

$$f_{jk} = \int \chi_j^*(\mathbf{r}) f(\mathbf{r}) \chi_k(\mathbf{r})\, d\mathbf{r} \,. \tag{9}$$

(The generalization to operators of higher order is obvious.)

In a model problem with a finite basis, it is possible that two or more operators that are distinct for the exact problem will have the same matrix in the model. They can thus not be distinguished in the model problem. In particular, a non-local operator like the kinetic energy or an SCF exchange operator may have a matrix which can be reproduced by a function, i.e. a local operator, in the model. In extreme cases (to be discussed later) one can define reasonable models in which *all* operators are equivalent to local operators.

2.5. Equidensity Orbitals and n Representability

2.5.1. Special Equidensity Orbitals

A particular choice of an orbital basis can be made which is specially suited to problems associated with a *given* density /5, 8, 9/. Suppose that $d(\mathbf{r}) = \rho(\mathbf{r})/n$ is a function which is nowhere negative and which is nonzero almost everywhere. (Typically, the density will vanish only at infinity.) It is then possible to construct a vector-valued function $\mathbf{F}(\mathbf{r})$ such that

$$\left| \frac{\partial(F_x, F_y, F_z)}{\partial(x, y, z)} \right| = 8\pi^3 d(\mathbf{r}) \,. \tag{10}$$

An explicit construction of such a function has been given /9/, although it does not lead to a computationally effective method. One can also imagine a basis set model in which this equation is approximately satisfied in some least squares sense. Given $\mathbf{F}$, a set of orthonormal functions is constructed as

$$\phi_{\mathbf{k}}(\mathbf{r}) = [d(\mathbf{r})]^{1/2} e^{i\mathbf{k}\cdot\mathbf{F}(\mathbf{r})} \tag{11}$$

where the components of $\mathbf{k}$ are all integers or all half-integers. This set of special equidensity orbitals (SEDOs) becomes complete in the limit as all integer (half-integer) values between $-\infty$ and ∞ are included for each component of $\mathbf{k}$.

2.5.2. n Representability

It is readily shown by the use of SEDOs that any density satisfying the positivity and normalization conditions is n representable /8/. From Eq. (11), $|\phi_{\mathbf{k}}|^2 = d$ for any $\mathbf{k}$, and thus any function having only SEDOs as natural orbitals will have density ρ. Such functions include any single determinant made up of SEDOs times spin functions, or any linear combination of such determinants where any two determinants in the expansion differ by at least two SEDO spinorbitals. Of particular interest for a singlet state, any expansion in terms of doubly occupied SEDOs will give ρ. More complicated and thus more flexible functions can also be constructed.

2.5.3. Absence of physical significance

Although the SEDOs are useful for formal manipulations, it seems unlikely that they can have any physical significance /10/. At the simplest level, one might imagine that there is a one-electron Schrödinger equation with some kind of effective potential of which the SEDOs are solutions. If the potential is assumed to be local, then it can be eliminated between two of the equations, giving a relationship between the two SEDOs involved. This becomes a differential equation that the density must satisfy, and such will be the case only in very rare, special circumstances. One could probably concoct an energy- and density-dependent nonlocal potential that would work for any specified, finite set of SEDOs, but it would be difficult to attach any physical significance to the result.

As an alternative, one might hope to express a set of SEDOs as a unitary transformation of some other set of orbitals that would not be equidensity but that could be, e.g., solutions of some effective one-electron equation. At the level of an independent particle model where the n-electron wavefunction will be a single determinant, such a transformation can be shown to be impossible except in special cases. This point will be discussed further below.

2.6. Negative Consequences of a Finite Model

Although the introduction of a finite basis model has many formal advantages and is often necessary in practice, it can have interesting negative consequences. These are unlikely to be of concern in a practical calculation but are nevertheless worthy of consideration.

2.6.1. Loss of n representability

It will be shown below that there are reasonable basis sets for which there is a one-to-one relationship between charge density matrices and densities. It is possible in at least some such bases to construct a matrix which has a negative eigenvalue and is thus not an n-representable reduced density matrix but which nevertheless corresponds to a density function that is nowhere negative /11/. This apparently reasonable density is thus not n representable.

This means that for models in which non-n-representable densities are possible, it is necessary to impose an n-representability constraint. Such a constraint can be explicitly stated when the densities are in one-to-one correspondence to charge density matrices because the conditions for n representability of a charge density matrix are know. In practice in such a case, since the density and density matrix have the same information content and involve the same number of independent parameters, it is probably more efficient to parameterize the density matrix in a way which guarantees its n representability and to use it rather than the density as the fundamental quantity.

2.6.2. Loss of the Hohenberg-Kohn theorem

It is also possible to find model problems for which the Hohenberg-Kohn theorem does not apply /11/. In particular, there can be Hamiltonian matrices which differ only by the matrix of a local potential, so that they correspond to systems with different local potentials, but which have the same ground state density.

For a one-particle system without spin, this situation can arise in a straightforward way. Start with some potential v_1, construct the Hamiltonian matrix with some finite basis set, and find its eigenfunctions. Then transform the basis

set to consist of these eigenfunctions, which will be denoted by $\psi_0, ..., \psi_{b-1}$ with ψ_0 the ground state. There are b products of the form $\psi_0^* \psi_j$, $j = 0, \ldots, b$. If a function Δv can be found which is expressible as some linear combination of all basis set products and is orthogonal to each of the products $\psi_0^* \psi_j$, then the Hamiltonian with potential $v_2 = v_1 + \Delta v$ will also have ψ_0 as an eigenvector, and Δv can be chosen (at least in most cases) so the ψ_0 is still the ground state eigenfunction of the new Hamiltonian.

For a many-electron system the starting point is the set of all products of eigenfunctions of the full CI Hamiltonian, $\Psi_0^* \Psi_J$. Integration over the coordinates of $n-1$ electrons and over the spin of the final electron gives a set of functions in the orbital product space. There will be many of them, and they may span the whole of $\mathcal{F}$. If they do not, however, a Δv orthogonal to all of them can be found and a new Hamiltonian matrix with the same ground state constructed, as above.

The conditions on a basis set which would assure a Hohenberg-Kohn theorem in the n-electron case must be very complicated, and have not been determined. A necessary condition is that the partially-integrated products $\Psi_0^* \Psi_J$ span $\mathcal{F}$. Fortunately, the Levy-Lieb functional remains well-defined and our analysis can proceed.

3. MATRIX AND DENSITY SPACES

In the preceding section, basis functions and model Hilbert spaces were discussed. We now consider spaces with matrices as elements.

3.1. Matrix Spaces $\mathcal{E}_p$

Consider the set of all linear operators from $\mathcal{H}_1$ to itself, i.e. the set of all one-electron operators restricted to the model problem. Every such operator can be represented by a $2b \times 2b$ matrix, and we introduce the space $\mathcal{E}_1$ of such matrices. For each Hilbert space $\mathcal{H}_p$ the operators restricted to the model problem correspond to $\binom{2b}{p} \times \binom{2b}{p}$ matrices in a space $\mathcal{E}_p$. For completeness, we include also the one-dimensional space $\mathcal{E}_0$ of "1×1 matrices." A scalar product can be

introduced for elements in each $\mathcal{E}_p$ as $(\mathbf{A},\mathbf{B}) = \mathrm{tr}\,\mathbf{A}^\dagger\mathbf{B}$, $\mathbf{A},\mathbf{B} \in \mathcal{E}_p$. Each of these spaces is thus a metric space and, in the finite-basis case, just a Euclidean vector space /12/.

It can be shown that each matrix space $\mathcal{E}_p$ can be decomposed into orthogonal subspaces

$$\mathcal{E}_p = \mathcal{E}_{p0} \oplus \mathcal{E}_{p1} \oplus \cdots \oplus \mathcal{E}_{pp} \tag{12}$$

The significance of these subspaces is in connection with the reduction map $\hat{L}_p^q$ from $\mathcal{E}_p$ to $\mathcal{E}_q$ [cf. Eq. (8)]. The index π of the subspace $\mathcal{E}_{p\pi}$ is called the reduction index. It has the following meaning: The operators or components of operators in $\mathcal{E}_{p\pi}$ are in one-to-one correspondence with operators or components of operators in $\mathcal{E}_{q\pi}$ for each q between p and π. If an attempt is made to reduce an operator in $\mathcal{E}_{p\pi}$ to $\mathcal{E}_q$ with $q < \pi$ the result is zero, i.e. the operator is annihilated by the attempted reduction. The CI $\mathbf{H}$ has components in $\mathcal{E}_{n0}$, $\mathcal{E}_{n1}$, and $\mathcal{E}_{n2}$, but not in subspaces with larger reduction indices.

The space $\mathcal{E}_1{}^0$ of spin-free one-electron operators contains $b \times b$ matrices. An important example of such an operator is the integral operator with the charge density matrix as kernel,

$$\gamma^0(\mathbf{r};\mathbf{r}') = \sum_{j,k} \gamma_{jk}\chi_j(\mathbf{r})\chi_k^*(\mathbf{r}')\,. \tag{13}$$

The Löwdin normalization $\mathrm{tr}\,\boldsymbol{\gamma} = n$ is assumed.

3.2. The Density Space

The "collapse" map $\hat{\delta}$ giving the density from the density matrix consists of setting $\mathbf{r}' = \mathbf{r}$ in the integral kernel /11/

$$\rho(\mathbf{r}) = \hat{\delta}\gamma^0 = \gamma^{(0)}(\mathbf{r};\mathbf{r}) = \sum_{jk} \gamma_{jk}\chi_j(\mathbf{r})\chi_k^*(\mathbf{r}) \tag{14}$$

All densities in the model problem can be expressed as linear combinations of products of the orbital basis functions. We denote by $\mathcal{F}$ the space spanned by these products.

Note however that even for orthonormal orbitals $\{\chi_j\}$, the products $\{\chi_j\chi_k^*\}$ will in general be linearly dependent. This linear dependency may be exact, for

basis functions of simple forms, or only approximate to within the numerical accuracy of a computer. Nevertheless, as the size of the basis is increased we must expect more and more linear dependencies among basis function products. It is because of this linear dependency that the density matrix contains more information than does the density /13/.

For any given basis set the linear dependencies are determined. It is then possible to decompose ${\mathcal{E}_1}^0$ into orthogonal subspaces defined with respect to the collapse map: ${\mathcal{E}_1}^0 = \mathcal{L} + \mathcal{K}$ where $\mathcal{L}$ consist of matrices associated with local operators and $\mathcal{K}$ is the null space of the collapse map. The dimension of $\mathcal{L}$ is the same as that of $\mathcal{F}$, and the elements of $\mathcal{L}$ are in one-to-one correspondence with those of $\mathcal{F}$. Despite the dependence of this decomposition on linear dependencies among basis set products, it can be shown that it really depends only on the orbital Hilbert space ${\mathcal{H}_1}^0$: it is not affected by a nonsingular transformation of the orbital basis /13/. It has also been shown that for a complete basis both $\mathcal{L}$ and $\mathcal{K}$ are infinite dimensional, and that in fact for any function $f(\mathbf{r})$ in $\mathcal{F}$ there is an infinite-dimensional subspace of ${\mathcal{E}_1}^0$ each element of which collapses to f. In particular, since the basis set model problem is the same as the physical problem when the basis is complete, any density can be obtained from an infinite number of charge density matrices. This was noted previously in connection with the n-representability of any density.

The SEDOs introduced earlier have particularly simple linear dependencies among their products. Each product is of the form

$$\phi_{\mathbf{j}}^*(\mathbf{r})\phi_{\mathbf{k}}(\mathbf{r}) = d(\mathbf{r})e^{i(\mathbf{k}-\mathbf{j})\cdot\mathbf{F}(\mathbf{r})} \,. \tag{15}$$

All products with the same value of $\mathbf{k}-\mathbf{j}$ will be not only linearly dependent, but identical. No set of orbitals arising from a physical problem is likely to have as many linear dependency conditions, relative to the number of orbitals in the set, as a SEDO set of comparable size. Since any unitary transformation of orbitals preserves the $\mathcal{L}, \mathcal{K}$ decomposition, the physical orbitals cannot be transformed to a SEDO set of the same or comparable size.

The characterization of the set of n-electron density matrices that give density ρ can now be explicitly given: ρ determines the component of γ in $\mathcal{L}$, call it $\gamma_{\mathcal{L}}$. Combining this with all elements of $\mathcal{K}$ defines a subspace of ${\mathcal{E}_1}^0$. The

intersection of this space with the set $\mathcal{P}_1^{(n)}$ of ensemble n-representable one matrices defines the subset $\mathcal{P}_\rho$ collapsing to ρ. The inverse of the reduction map applied to $\mathcal{P}_\rho$ defines a set in $\mathcal{E}_{n0} \oplus \mathcal{E}_{n1}$. Combine this set with $\mathcal{E}_{n2} \oplus \ldots \oplus \mathcal{E}_{nn}$ and then intersect the resulting set with the set $\mathcal{P}_n$ of all positive, Hermitian, trace 1 matrices in $\mathcal{E}_n$. This intersection is the required set.

3.3. Linearly Independent Product Bases

3.3.1. Definition of a LIP Basis

If a basis set of simple analytic form such as trigonometric functions or harmonic oscillator eigenfunctions is used, then exact linear dependencies among products arise very quickly. The number of linear dependency conditions can easily be found by simple counting arguments. Consider for example the oscillator functions $\psi_n = C_n H_n(x) e^{-x^2/2}$ for $0 \leq n \leq n_{\max}$. Each product $\psi_{n_1}\psi_{n_2}$ will be of the form $P_{n_1 n_2}(x)e^{-x^2}$ where $P_{n_1 n_2}$ is a polynomial of degree $n_1 + n_2$. The number of products possible is $n_{\max}(n_{\max}+1)/2$, but there are to most $2n_{\max}+1$ linearly independent polynomials of degree not exceeding $2n_{\max}$.

With basis sets more typical of those used in practical calculations, such as Cartesian Gaussian functions on various atomic centers, it is very unlikely that exact linear dependency conditions will arise. Simple examples show, however, that near linear dependencies can quickly become a problem. An attempt to orthogonalize the set of product functions, or to solve an generalized eigenvalue problem with the product functions as basis, will fail because the overlap matrix will have eigenvalues very close to zero.

If one is willing to sacrifice some accuracy, or if the number of electrons is not too large and the basis is chosen with great care, then a model problem can be defined for which all basis function products are linearly independent. Such a basis will be referred to as LIP /13/.

3.3.2. Consequences of a LIP basis

Any one-electron operator $\hat{f}$ is defined in the model problem by its matrix elements f_{jk}. Now consider a function ϕ defined as a linear combination of basis orbital products, $\phi(\mathbf{r}) = \sum_{lm} \phi_{lm}\chi_l(\mathbf{r})\chi_m^*(\mathbf{r})$. Can the expansion coefficients $\{\phi_{lm}\}$ be chosen so that the matrix of ϕ considered as a potential is the same as that of $\hat{f}$? The answer is always "yes" for a LIP basis. The requirement is that

$$f_{jk} = \sum_{lm} \Delta_{jk,lm}\phi_{lm} \tag{16}$$

where

$$\begin{aligned} \Delta_{jk,lm} &= \int \chi_j^*(\mathbf{r})[\chi_l^*(\mathbf{r})\chi_m(\mathbf{r})]\chi_k(\mathbf{r})\,d\mathbf{r} \\ &= \int \left[\chi_j^*(\mathbf{r})\chi_k(\mathbf{r})\right][\chi_l^*(\mathbf{r})\chi_m(\mathbf{r})]^*\,d\mathbf{r} \end{aligned} \tag{17}$$

is just the transpose of the overlap matrix for the orbital products. Since this matrix is nonsingular for a LIP basis, it can be inverted and a ϕ yielding $\mathbf{f}$ can be found. This means that in a LIP basis any operator is equivalent to a local potential.

If a LIP-basis model can be constructed which gives a good approximation for a problem of interest, then it will have a number of interesting properties. Exact local expressions can be found for the kinetic energy and SCF exchange energy. It cannot be assumed, however, that these expressions will be similar to commonly-used local density functional approximations. Their significance thus remains open to question. For example, the kinetic energy functional is $T[\rho] = \int t(\mathbf{r})\rho(\mathbf{r})\,d\mathbf{r}$ where t is independent of ρ but depends on the basis. A one-electron density matrix, and thus possibly a single-determinant wavefunction, can be determined by experimental measurements sensitive only to the density, such as x-ray diffraction, only within the context of a LIP model.

An important question now arises: How good can a LIP model be? No complete basis can be LIP /15/, so we cannot hope for exact results even as a limit. On the other hand, many of the basis sets that have been used for calculations of reasonable quality are LIP. From a formal point of view, there must be an optimum LIP basis for any problem. No investigation of the precise characterization of such a basis or of an example has yet been reported.

4. ENERGY CONTRIBUTIONS

The energy of the electronic system, and thus the energy functional whose existence is established by the Hohenberg-Kohn theorem, are usually divided into sums of contributions. This is always possible so long as each new piece is rigorously defined and as the final piece is defined to be the total minus the sum of the other pieces thus far defined. Such a division is useful provided that the individual pieces as rigorously defined correspond closely to the different physical contributions to the energy they are intended to describe. The original and "cleanest" separation is that of the interaction with the external potential, leaving a functional $F[\rho]$, as in Eq. (1). Several other contributions can be separated with varying degrees of success. It should be recalled that the important criterion in the present discussion is formal justification rather than practical utility.

4.1. Coulomb Energy

It is certainly possible to define a Coulomb energy contribution in a straightforward and useful way. Let

$$U_{\mathrm{Coul}}[\rho] = \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\mathbf{r}d\mathbf{r}' \,. \tag{18}$$

This is immediately recognizable as the classical Coulomb energy, and is a straightforward local functional of the density. The principal problem with separation of this term is that includes a self-energy contribution that must be removed elsewhere. This point will be further discussed below.

4.2. Kinetic Energy

The separation of kinetic from potential energy contributions is certainly attractive, but is surprisingly difficult to do in a well-defined way. One reason is that there is no simple kinetic energy functional of the density. The kinetic energy is inherently a nonlocal property, and expansions in terms of gradients of the density lose their attractiveness after a few terms.

The Kohn-Sham approach uses the kinetic energy of a system of noninteracting electrons having the same density as that specified /14/. It is not as clear

as one might wish just what this really means, or whether the resultant kinetic energy is really the total kinetic energy. A well-defined starting point is to use the constrained search algorithm for the kinetic energy only. Having introduced a basis set, we define

$$T[\rho] = \min_{\gamma \in \mathcal{P}_{1\rho}} \mathrm{Tr}\, \mathbf{t}\boldsymbol{\gamma} \tag{19}$$

where $\mathbf{t}$ is the matrix of one-electron kinetic energy operator, $\boldsymbol{\gamma}$ is the charge density matrix, and $\mathcal{P}_{1\rho}$ is the set of all ensemble n-representable $\boldsymbol{\gamma}$'s that yield the density ρ.

If a SEDO basis is used the characterization of $\mathcal{P}_{1\rho}$ is very simple /15/. Consider a truncated set of SEDOs in which each component of $\mathbf{k}$ is restricted to the range $-\nu$ to ν. The restricted kinetic energy functional $T_\nu[\rho]$ is an upper bound to $T[\rho]$ which becomes tighter as ν is increased. It can be shown that

$$t_W[\rho] + \sum_{\mu}^{(n)} \kappa_\mu \leq T_\nu[\rho] \leq t_W[\rho] + \sum_{\mathbf{k}}^{(n)} K_{\mathbf{k},\mathbf{k}} \tag{20}$$

where

$$t_W[\rho] = \frac{1}{8} \int \frac{(\nabla\rho)^2}{\rho}\, d\mathbf{r}\,, \tag{21}$$

$\mathbf{K}$ is a matrix with elements

$$K_{\mathbf{jk}} = \frac{1}{3} \int \rho \nabla(\mathbf{j}\cdot\mathbf{f})\cdot\nabla(\mathbf{k}\cdot\mathbf{f}) \exp[i(\mathbf{k}-\mathbf{j})\cdot\mathbf{f}]\, d\mathbf{r}\,, \tag{22}$$

and κ_μ is an eigenvalue of $\mathbf{K}$. The sums extend over the n smallest eigenvalues or the n smallest diagonal elements. The actual value of $T_\nu[\rho]$ can in principle be obtained numerically for any ν.

A more general analysis of the constrained search kinetic energy functional shows that it is consistent with the Kohn-Sham method /16/. The kinetic energy for a given density will be that of a single determinant function where the orbitals in the determinant are solutions of a one-electron Hamiltonian with an appropriate effective potential.

It has been pointed out that an exact treatment of the kinetic energy is possible if the one-electron density matrix rather than the density is taken as the fundamental quantity. An alternative is to assume knowledge of the momentum density $\pi(\mathbf{p})$, since $T = \frac{1}{2}\int p^2\pi(\mathbf{p})\, d\mathbf{p}$. Potential energy functionals of the

momentum density are even worse that the kinetic energy functional of the coordinate density, so a purely momentum DFT is not very attractive, but a theory based on functionals of $\pi(\mathbf{p})$ and $\rho(\mathbf{r})$ is possible. These two densities together require less information than the density matrix, and have the added advantage that they are in principle observable.

4.3. Exchange-Correlation Energy

The meaning of the term "correlation" in electronic quantum mechanics is not the same as it is in probability theory. A large part of what would be correlation is separated out as an "exchange" effect. While not restricted to SCF theory, these ideas have their cleanest expression there, and SCF theory is precisely that level of approximation where everything is determined by the one matrix.

In a truly uncorrelated system, the two-electron density would be just a product of one-electron densities: $\mathsf{P}(\mathbf{r}_1, \mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)$. To examine exchange, one must consider the density matrices rather than just densities. For a two-electron density matrix in the Löwdin normalization $\mathrm{tr}\,\boldsymbol{\Gamma} = n(n-1)/2$, one can think of defining a part of Γ in terms of γ as /17, 18/

$$\Gamma_{cx}(1,2;1',2') = \frac{1}{2}\left[\gamma(1;1')\gamma(2;2') - \gamma(1;2')\gamma(2;1')\right] . \tag{23}$$

The subscript cx refers to the obvious interpretation as Coulomb and exchange contributions. It is well known that for a state described by a single determinant this is exact for all of $\boldsymbol{\Gamma}$. It is less well known that $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_{cx}$ *requires* a single determinant state. If the consistency condition $\gamma(1;1') = \frac{2}{n-1}\int \Gamma_{cx}(1,2;1',2)\,d2$ is imposed it is found to require $\boldsymbol{\gamma}^2 = \boldsymbol{\gamma}$. This idempotency arises only for one-matrices arising from single determinants. The spin-polarized form of this relationship involves "Coulomb" interactions among all electrons but "exchange" interactions only among electrons of the same spin. The consistency condition then requires that the spin-up and spin-down one-matrices each be idempotent and again implies a (spin-polarized) single determinant.

The conclusion to be drawn is that if an attempt is made to separate an exchange functional in a model more general than SCF, then the left-over part will not be pure (correction to SCF) correlation, but will also include a correction

to the exchange contribution. It may still be useful in some contexts to make an approximate separation of these effects because the exchange piece is such a large part of the total.

It is clear from Eq. (23) that the probability of finding two electrons with the same spin at the same place is zero. This will be true for any model that takes proper account of the Pauli principle. When SCF theory is developed in terms of the Fock operator eigenfunctions, the energy expression involves a sum over all orbitals *except* the orbital that the operator is acting on. This sum can be extended to include that orbital as well because Coulomb and exchange contributions cancel and the total expression is the same whether it is included or not. If Coulomb and exchange contributions are separated, then this extra term will be responsible for the apparent "self energy" contribution in the Coulomb terms.

5. CONCLUSIONS

It has been noted here that basis sets are often appropriate in density functional theory, for formal or practical reasons or both. In the DFT associated with a basis-set model, linear dependency conditions among basis set products play a key role.

The Levy-Lieb-Valone constrained search algorithm defines an energy functional that remains valid in basis set models where the Hohenberg-Kohn theorem does not apply and can be explicitly stated in terms of the collapse map and one-matrix n-representability conditions. Only a limited set of densities are possible in a basis-set model, and some of them may fail to be n representable. For any reasonable density, however, one can choose an appropriate (potentially complete) basis in terms of which that density is n representable.

A constrained search kinetic energy functional can be stated explicitly. Alternatively, the kinetic energy can be expressed exactly in terms of the one-matrix or in terms of the momentum density. A joint functional $E[\rho, \pi]$ would involve coordinate and momentum densities, both of which are experimentally observable. Both densities are determined by the one-matrix and they are linked by

consistency conditions. Relevant to this is the renewed interest in a phase-space description, whether in terms of the Wigner function /19, 20/, which has ρ and π as marginals, or in terms of the Husimi function /21, 22/, which is a probability density for finding an electron of the system in a Gaussian wavepacket (coherent state) centered at a specified point in phase space.

In the model defined by a LIP basis, exact local functionals exist for the kinetic energy and SCF exchange energy. The one-matrix and momentum density are determined by ρ in a LIP basis. Such a basis cannot become complete and experience indicating how good a description of a physical system can be provided by a LIP basis remains to be developed.

It has been noted that an exchange contribution to the energy can be separated in a consistent way only for an SCF-like treatment. Also, as is well known, self energy problems are associated with the separation of a Coulomb contribution from the rest of the two-electron energy. A basis set analysis does not eliminate these problems. While the set of two-electron density matrices associated with a given density is easily characterized, usable n-representability conditions on the two-matrix are not known.

Despite these limitations, the use of basis set models, especially with particular choices of basis orbitals, can make major contributions to our understanding of density functional theory and of its relationship to density matrix theory.

ACKNOWLEDGMENT

Some of the results reported here were obtained with support from the National Science Foundation through grants CHE 8201909 and CHE 8519723.

REFERENCES

1. P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).

2. A. J. Coleman, *Rev. Mod. Phys.* **35**, 668 (1963).

3. This term was coined by E. G. Larsen in 1975.

4. M. Levy, *Proc. Nat. Acad. Sci. USA* **76**, 6062 (1979).

5. E. Lieb, *Int. J. Quantum Chem.* **24**, 243 (1983).

6. S. M. Valone, *J. Chem. Phys.* **73**, 1344 (1980).

7. H. Kummer, *J. Math. Phys.* **8**, 2063 (1967).

8. J. E. Harriman, *Phys. Rev. A* **24**, 680 (1981).

9. G. Zumbach and K. Maschke, *Phys. Rev. A* **28**, 544; **29** 1585(E) (1983).

10. J. F. Capitani, B. Chang, and J. E. Harriman, *J. Chem. Phys.* **81**, 349 (1984).

11. J. E. Harriman, *Phys. Rev. A* **27**, 632 (1983).

12. For a discussion and references see M. E. Casida and J. E. Harriman, *Int. J. Quantum Chem.* **30**, 161 (1986).

13. J. E. Harriman, *Phys. Rev. A* **34**, 29 (1986).

14. W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).

15. J. E. Harriman, *J. Chem. Phys.* **83**, 6283 (1985).

16. W. Yang and J. E. Harriman, *J. Chem. Phys.* **84**, 3320 (1986).

17. P. A. M. Dirac, *Proc. Cambridge Philos. Soc.* **27**, 240 (1931).

18. P. O. Löwdin, *Phys. Rev.* **97**, 1474 (1955).

19. E. Wigner, *Phys. Rev.* **40**, 749 (1932).

20. For a recent discussion and additional references, see M. Hillery, R. F. O'Connell, M. O. Skully, and E. P. Wigner, *Phys. Repts.* **106**, 121 (1984).

21. K. Husimi, *Proc. Phys.-Math. Soc. Japan* **22**, 264 (1940).

22. J. E. Harriman, *J. Chem. Phys.* **88**, 6399 (1988).

DENSITY MATRIX FOUNDATIONS OF DENSITY FUNCTIONAL THEORY: THE IMPORTANCE OF PURE-STATE N-REPRESENTABILITY IN THE DERIVATION OF EXTENDED KOHN-SHAM EQUATIONS

Eduardo V. Ludeña

Química, Instituto Venezolano de Investigaciones Científicas, Caracas 1020-A, Venezuela.

and

Jaime Keller

División de Ciencias Básicas,
Facultad de Química, Universidad Nacional Autónoma de México,
Ciudad Universitaria, 70-528, 04510 México, D.F., México.

1. INTRODUCTION.

We discuss in this chapter, within the context of a density matrix approach based on the 1-matrix, where pure-state N-representability conditions are explicitly incorporated, how the Schrödinger equation may be exactly reduced to a set of 2m coupled equations (m is the number of single particle orbitals) which give solutions for both the ground state and excited states.

These coupled equations, given by Nguyen-Dang et al. via a realization of the two-step variational procedure introduced by Levy and Lieb, are shown to yield, under appropriate simplifications the usual Kohn-Sham equations. The relevance of these novel equations, in relation to molecular physics and condensed matter theory calculations, is discussed.

The basic aspiration of much of the work done in density

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

functional theory and density matrix theory /1/ has been to set up a formalism totally equivalent to the Schrödinger equation, for pure states or to the Liouville equation, for ensembles which, nevertheless, bypasses the use of N-particle wavefunctions or N-matrices, respectively /2/. Due to the particular nature of the Hamiltonian for Coulomb systems, where at the most one has to deal with two-particle interactions, we could, in principle, reduce the many-body problem into an equivalent problem totally cast in terms of the 2-matrix, $D^2(1,2;1',2')$ an object which regardless of the number of particles of the system always depends upon only 12 spatial coordinates. This represents a considerable simplification if we consider that the N-particle wave function for systems of interest in chemistry, materials science or biology depends upon 3N spatial coordinates, where N can easily attain the value of thousands. But unfortunately, the construction of an alternative formalism based on the 2-matrix has met with a difficulty which until now is unsurmountable: the pure-state N-representability problem /3/. This problem has to do with the necessary and sufficient conditions which must be satisfied by the 2-matrix in order that the variation of the energy expression, solely based on $D^2(1,2;1',2')$, yield the exact ground-state energy of the N-particle system.

The difficulties which one encounters in the case of the 2- matrix persist when one attempts to reformulate this problem on the 1-matrix $D^1(1,1')$ or, even simpler, in terms of its diagonal part, the one-particle density $\rho(1) = D^1(1,1)$. The pure-state N-representability problem for the 1-matrix is still unsolved /4/. This, plus the fact that the explicit expression for the energy as a functional of the 1-matrix is unknown has barred the way to a direct construction of a variational principle based on $D^1(1,1')$. However, when the 1-matrix does not refer to pure states, the N-representability conditions are known /5/ and hence, in this context, a quantum theory for ensembles may be developed. The only obstacle to

such an undertaking still is the lack of an explicit functional form for the energy in terms of the 1-matrix.

An alternative approach to the variational problem for pure-state 1-matrices, where use is made of the known N-representability conditions for ensemble 1-matrices, has been advanced by Nguyen-Dang, Ludeña and Tal /6/. This seemingly unusual approach may be justified by resorting to the two-step procedure of Levy /7/ and Lieb /8/, which is founded, in turn, on the Hohenberg-Kohn theorems /9/.

It is the purpose of the present work to extend the results of Nguyen-Dang et al. /6/ in order to show that the variational principle not only leads to coupled equations for the ground state of the system, but that it also provides us with a first description of excited states. Furthermore, we show that the usual Kohn-Sham equations /10/ can be regarded as an approximation to these exact orbital equations. The alternative perspective whereby these coupled equations may be viewed as extended Kohn- Sham equations where the symmetry of each state has been explicitly introduced, is also discussed.

In Section 2, we succinctly review the Hohenberg-Kohn theorems, and their relation to the two-step variational principle of Levy /7/ and Lieb /8/. In Section 3, we describe the constrained variational procedure which is used in the determination of the auxiliary universal functional and analyze in this respect the importance of N-representability conditions. In Section 4 we discuss the connections between the system of coupled orbital equations which result from this formalism and those of Kohn-Sham /10/. In Section 5 we discuss the problem of multiconfiguration in terms of $D^1(1,1')$ and finally in Section 6, as an example, we discuss the relation of this approach and a particular type of problem in condensed matter physics.

2. THE HOHENBERG-KOHN THEOREMS AND THE TWO-STEP VARIATIONAL PROCEDURE.

Consider a Hamiltonian operator $\hat{H}_v$ defined as the sum of the kinetic energy operator $\hat{T}$, the electron-electron interaction operator $\hat{V}_{ee}$ and an external potential operator $\hat{V}_{ext}$:

$$\hat{H}_v = \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} = \hat{H}_o + \sum_{i=1}^{N} V(\vec{r}_i) \qquad (1)$$

where $V(\vec{r}_i)$ is the i-th component of the external potential. We shall refer to this Hamiltonian as belonging to the H_o-family of Hamiltonians. Ordinarily, this external potential describes the nuclear-electron interaction.

Let us assume that $\hat{H}_v$ has a non-degenerate ground state Ψ_o^v which is a solution to Schrödinger's equation $\hat{H}_v \Psi_o^v = E_o^v \Psi_o^v$. Then, we may define the following set of external potentials (which need not necessarily describe nuclear-electron interactions):

$$V_N \equiv \left\{ V(\vec{r}) \,\middle|\, \hat{H}_v = \hat{H}_o + \sum_{i=1}^{N} V(\vec{r}_i) \quad \begin{array}{l}\text{possesses a non-degene-}\\ \text{rate ground state}\end{array} \right\} \qquad (2)$$

It is clear from this expression, however, that not any arbitrary external potential may be used in Eq. (1). Associated to each ground-state wavefunction Ψ_o^v we have a ground-state density $\rho_o^v(\vec{r})$ defined by (here $\int$ includes spin sums)

$$\rho_o^v(\vec{r}_1) = \sum_{\sigma_1} \int d2 \ldots \int dN \; \Psi_o^{v*}(1,2,\ldots,N) \Psi_o^v(1,2,\ldots,N) \qquad (3)$$

consider now the set of one-particle densities

$$A_v^N = \left\{ \rho(\vec{r}) \,\middle|\, \text{such that } \rho(\vec{r}) \text{ comes from a ground-state wave-function corresponding to an } H_o \text{ -family of Hamiltonians with } V(r) \in V_N \right\} \qquad (4)$$

equation (4) is the general expression for what is known as the representability condition on a density.

The first Hohenberg-Kohn theorem establishes a one-to-one correspondence between $\rho(r) \in A_v^N$ and $V(r) \in V_N$ which, symbolically may be put as $\rho_o^v(r) \longleftrightarrow V(r)$. The above correspondence indicates that for a given density $\rho_o^v(\vec{r}) \in A_v^N$, we can univocally obtain the external potential V(r).

Through this inverse mapping, the exact ground-state energy E_o^v may be determined from the sole knowledge of $\rho(r)$. In other words, $E_o^v(r)$ is a functional of the ground-state density

$$E_o^v = E_o^v\left[\rho_o^v(\vec{r})\right] \tag{5}$$

equation (5) embodies the statement of the second Hohenberg-Kohn theorem.

There are the following points which we would like to emphasize here. In the first place, the Hohenberg-Kohn theorems refer specifically to ground state densities corresponding to the H_o-family of Hamiltonians and not to any arbitrary density. In the second place, the functional expression given by Eq. (5) is not an explicit one. In fact, $E_o^v[\rho_o^v]$ must be interpreted in terms of the ground-state wavefunction as follows

$$E_o^v\left[\rho_o^v(\vec{r})\right] \equiv \langle\Psi_o^v|\hat{T} + \hat{V}_{ee}|\Psi_o^v\rangle + \int d\vec{r}\rho_o^v(\vec{r})V(\vec{r}), \tag{6}$$

this means that the second Hohenberg-Kohn theorem does not tell us how to express either the kinetic energy or the electron-electron interaction energy in *bona fide* functionals of the ground-state density $\rho_o^v(\vec{r})$. In this sense, it is just an existence theorem valid only at the exact ground state level. (Remember that $E_o^v\left[\rho_o^v\right]$ not being known we could resort to $E_o^v\left[D'(1,1')\right]$ as a more practical approach, without invalidating (5)).

In order to further clarify the above ideas, consider the inequality

$$\langle\Psi_\rho|H_o + \sum_{i=1}^{N} V(\vec{r}_i)|\Psi_\rho\rangle \geq E_o^v \; ; \quad \Psi_\rho \rightarrow \rho(\vec{r}) \in N_{HK}^N \quad \text{and} \quad \Psi_\rho \in \mathcal{L}_N \tag{7}$$

where the Hohenberg-Kohn set N_{HK}^{N} of one-particle densities is defined by

$$N_{HK}^{N} = \left\{ \rho(\vec{r}) \,|\, \rho(\vec{r}) \geq 0 \ , \quad \int d\vec{r}\rho(\vec{r}) = N \right\} \ , \tag{8}$$

and where $\mathcal{L}_N$ is a subset of the antisymmetric N-particle Hilbert space $L^2_\sigma(R^{3N})$ and is given by

$$\mathcal{L}_N = \left\{ \Psi \,|\, \Psi \in L^2_\sigma(R^{3N}) \ ; - \langle \Psi | \nabla^2 | \Psi \rangle < \infty \ , \ \langle \Psi | \Psi \rangle = 1 \right\} \tag{9}$$

clearly, the equality in Eq. (7) is attained when $\Psi_\rho = \Psi_{\rho_o^v} \equiv \Psi_o^v$. Notice that the exact ground-state density $\rho_o^v(\vec{r})$ belongs to N_{HK}^N and also that A_v^N is a proper subset of N_{HK}^N :

$$A_v^N \subset N_{HK}^N \ . \tag{10}$$

For a one-particle density which satisfies the condition $\rho(\vec{r}) \notin A_v^N$, $\rho(\vec{r}) \in N_{HK}^N$. i.e. $\rho(\vec{r}) \in N_{HK}^N \setminus A_v^N$ the first Hohenberg-Kohn theorem does not guarantee a one-to-one correspondence between $\rho(r)$ and $V(r)$. In fact, it has been shown that there exist densities $\rho(\vec{r}) \in N_{HK}^N$ for which it is not possible to define a $V(r)$. Namely, densities which are not V-representable /11/. Now, if there is not a one-to-one correspondence, then there is not an inverse mapping and hence one cannot express the exact state energy as a functional of $\rho(r)$.

A variational principle for the energy, based on the one-particle density, must be restricted, according to the Hohenberg-Kohn theorem, to those densities which belong to the set of V-representable densities A_v^N. Such a variational principle is given, for example, by

$$E_o^v = \inf\left\{ \langle \Psi_\rho | \hat{H}_o | \Psi_\rho \rangle + \int d\vec{r}\rho(\vec{r})V(\vec{r}) \ ; \ \Psi_\rho \longrightarrow \rho(\vec{r}) \in A_v^N \ , \ \Psi_\rho \in \mathcal{L}_N \right\} \tag{11}$$

Notice that since $\rho(\vec{r})$ is restricted to belong to the set A_v^N, the N-particle wavefunctions $\Psi_\rho \in \mathcal{L}_N$ are, in fact, also restricted to be wavefunctions which yield V-representable one-electron densities. The problem with this restriction is that the necessary and sufficient conditions which describe A_v^N remain unknown.

Let us comment here in passing that a correct interpretation of the energy functional in Eq.(11) has been given by Nakatsuji and Parr /12/ by assuming that the infimum is searched along a continuous path leading from some reference external potential $V_{ref}(r) \in V_N$ to the actual potential $V(r)) \in V_N$. In this way, the functional is only defined for ground-state wavefunctions corresponding to external potentials along these continuous paths.

The two-step constrained variation of Levy /7/ and Lieb /8/ avoids the need to rely on the unknown set A_v^N. This is done by defining a universal functional $F_{LL}[\rho(\vec{r})]$ (LL stand for Levy-Lieb)as follows

$$F_{LL}[\rho(\vec{r})] = \inf\left\{ < \Psi_\rho | \hat{H}_o | \Psi_\rho > \; ; \; \Psi_\rho \longrightarrow \rho(\vec{r}) \in N_{HKL}^N \; , \; \Psi_\rho \in \mathcal{L}_N \right\} \quad (12)$$

where $\rho(r)$ is a <u>given</u> density belonging to the Hohenberg-Kohn-Lieb set N_{HKL}^N which is defined by

$$N_{HKL}^N = \left\{ \rho(\vec{r}) \,|\, \rho(\vec{r}) \geq 0, \; \int d\vec{r}\rho(\vec{r}) = N, \; \nabla\rho \in L^2(R^3), \right.$$

$$\left. \int d\vec{r}\left[\nabla\rho(\vec{r})^{1K}\right]^2 < \infty \right\} \quad (13)$$

The energy is then calculated by means of the second variational step

$$E_o^v = \inf\left\{ F_{LL}[\rho(\vec{r})] + \int d\vec{r}\rho(\vec{r})V(\vec{r}) \; ; \; \rho(\vec{r}) \in N_{HKL}^N \right\} \quad (14)$$

Let us notice that N_{HKL}^N is contained in N_{HK}^N. Hence, we have the following inclusions

$$A_v^N \subset N_{HKL}^N \subset N_{HK}^N \tag{15}$$

Eq. (12) implies a constrained variation, where at all stages the condition $\Psi_\rho \longrightarrow \rho(\vec{r}) \in N_{HKL}^N$ must be observed. This means in practice that if Ψ_ρ is represented in some general form, such as for example a configuration interaction (CI) expansion with M Slater determinants Φ_I constructed from an orbital set $\left\{\phi_{i_k}\right\}$ /13/

$$\Psi_\rho(1,\ldots,N) = \sum_{I=1}^{M} C_I \Phi_I(1,\ldots,N) \; ; \; I \equiv \left\{i_1 \, , \, i_2 < \ldots < i_N\right\} \tag{16}$$

then, this condition may be implemented by requiring that

$$\left(\rho_{\Psi_\rho}^*\right)_{ij} = \sum_{I=1}^{M} \sum_{J=1}^{M} C_I C_J^* \Delta_{IJ}^{ij} = \rho_{ij} \tag{17}$$

this follows from the fact that if we introduce the CI wavefunction given by Eq. (16) into Eq. (3), then the one-particle density (expressed in terms of the orbital set $\left\{\phi_{i_k}\right\}$) becomes

$$\rho_{\Psi_\rho}^*(1) = \sum_{I=1}^{M} \sum_{J=1}^{M} \left(\rho_{\Psi_\rho}^*\right)_{ij} \phi_i(1)\phi_j(1) \; ; \quad i = i_K \, , \quad j = j_\ell \tag{18}$$

$$\Delta_{IJ}^{ij} = \left\{(-1)^{K-\ell} \text{ if } \{J-i\} = \{J-j\} \; ; \; \text{zero otherwise}\right\}$$

for a given density (expressed in terms of the same orbital basis) the values of {CI} are fixed (the only conditions which $\rho(1)$ = must satisfy are those that ensure that it belongs to N_{HKL}^N). This means that at all stages of the variation there are some restrictions on the values which the CI coefficients may take. These conditions enter directly into the characterization of the universal function $F_{LL}[\rho(\vec{r})]$ /6/,/13/.

The crucial point which we would like to stress is that the requirement that the N-particle wavefunction correspond to a pure state is a basic ingredient of the first variational step leading to the universal functional $F_{LL}[\rho(\vec{r})]$. Since in the second step the one-particle density is varied until the infimum of the right

hand side of Eq. (14) is attained, it is clear that at the extreme point of variation, the CI wavefunction would become the exact ground-state wavefunction Ψ_o^v /13/. In other words, the essential condition which makes it possible to reach the exact ground-state energy by means of this two-step variational procedure is the pure-state-representability of the one-particle density.

Because in Eqs. (12) and (14), $\rho(\vec{r}) \in N_{HKL}^N$ is apparently the only requirement on $\rho(r)$, and since for any one-particle density $\rho(\vec{r}) \in N_{HKL}^N$ it is always possible to construct a single Slater determinant from which this density may be obtained, it has been assumed that $\rho(\vec{r}) \in N_{HKL}^N$ already implies that the condition of pure-state N-representability is satisfied. But as we have argued above, these conditions must be incorporated into the variational problem at the level of the constrained variation whereby $F_{LL}[\rho(\vec{r})]$ is constructed.

Furthermore, the fact that $\rho(\vec{r})$ may be derivable from a single Slater determinant does not guarantee that the infimum value of the right hand side of Eq. (14) yield the exact ground-state energy E_o^v. For this reason, any arbitrary or approximate universal function $\tilde{F}_{LL}[\rho(\vec{r})]$ may not be used when implementing the second variational step. A non-N-representable functional would invalidate the variational principle so that depending on the nature of the approximation, the right hand side of Eq. (14) may yield values above or even below E_o^v.

3. BUILT-IN PURE-STATE N-REPRESENTABILITY CONDITIONS FOR $D^1(1,1')$.

There are two basic reasons for choosing the 1-matrix $D^1(1,1')$ instead of $\rho(\vec{r})$ as the basic quantity in an "orbital" version of density functional theory. The first, is that the kinetic energy is an exact functional of $D^1(1,1')$. The second, that the ensemble N-representability conditions on $D^1(1,1')$ are

known. Furthermore, if one desires to derive one-particle orbital equations, then the most reasonable starting point is a variational principle for $D^1(1,1')$ /14/.

In order to bypass from the outset the V-representability problem, one may use an approach similar to that of Levy /7/ and Lieb /8/ and set up the two-step variational procedure for $D^1(1,1')$. This has been done by Nguyen-Dang et al. /6/ in the following way. The universal functional is defined by

$$F[D^1] = \inf \left\{ \mathrm{Tr}\ [\hat{H}_o D^N] \ ;\ D^N \in P_N \ ,\ D^N \longrightarrow D^1 \in P_N^1 \subset E_N^1 \right\} \quad (19)$$

and then the energy is obtained by means of

$$E_V[D^1] = \inf \left\{ F[D^1] + \int d\vec{r}\ V(\vec{r})\ D^1(r,r\,) \ ;\ D^1 \in E_N^1 \right\} \quad (20)$$

in the above expression, P_N is the set of pure-state N-matrices D^N and is given by

$$P_N = \left\{ D^N \,|\, D^N \equiv |\Psi\rangle\langle\Psi|,\ D^N \geq 0,\quad (D^N)^2 = D^N \ ,\ (D^N)^+ = D^N \ ,\right.$$
$$\left. \mathrm{Tr}\ D^N = 1,\ \Psi \in \mathcal{L}_N \right\} \quad (21)$$

P_N^1 is the corresponding set of pure-state N-representable 1-matrices $D^1(1,1')$. The necessary and sufficient condition which characterize P_N^1 remain unknown. This is not the case for E_N^1, the set of ensemble representable 1-matrices, whose complete characterization is given by the following conditions /5/.

$$E_N^1 = \left\{ D^1(1,1') \,|\, D^1(1,1') = \sum_{i=1} \lambda_i \psi_i^*(1)\psi_i(1') \ ;\ 0 \leq \lambda_i \leq 1, \right.$$
$$\left. \sum_{i=1} \lambda_i = N \right\} \quad (22)$$

where the λ_i's are the occupation numbers and the $\psi_i(1)$'s, the natural spin orbitals of $D^1(1,1')$ /15/.

The constrained variation indicated in Eq.(20) is carried out using the auxiliary functional /6,13/

$$E[D^N] = Tr[\hat{H}_o D^N] - \mathcal{E}\,(TrD^1 - 1) - Tr_1\left(\underline{\alpha}\,(L_N^1 D^N - D^1)\right) \qquad (23)$$

where $\mathcal{E}$ is the Lagrange multiplier which includes the normalization condition in $D^1(1,1')$ and $\underline{\alpha}$ is the Lagrange mulplier matrix which guarantees that

$$D_\Psi^1(1,1') = L_N^1 D_\Psi^N = \int d2 \ldots \int dN\, \Psi^*(1,2,\ldots,N)\, \Psi\,(1',2,\ldots,N) \qquad (24)$$

reduces to a given 1-matrix $D^1(1,1') \in E_N^1$. The above restriction is equivalent (in a CI representation of the N-matrix) to Eq. (17). Notice that at this point, or at any other point of the variation, all that is required from the $D^1(1,1')$ is that they belong to E_N^1. Since the variation must be restricted to pure-state N-matrices, the additional condition (idempotency) is also imposed

$$\delta D^N = D^N\, \delta\, D^N + (\delta D^N)\, D^N \qquad (25)$$

carrying out the first variational step one gets /6/ the universal

$$F[D^1] = Tr\left[\hat{H}_o D_o^N\right] = Tr_1[\alpha' D'] = \int d1\; \alpha'(1) D^1(1,1')\big|_{1' \to 1} \qquad (26)$$

where D_o^N. is the N-matrix at the extreme and the one particle operator $\underline{\alpha}$ is given in a matrix representation by

$$\alpha'_{ij} = \alpha_{ij} - \frac{\mathcal{E}}{N} \qquad (27)$$

the effective one-particle operator is essentially, therefore, just the Lagrange multiplier corresponding to Eq. (23). Defining

$$H_v[D^1] = Tr_1[(\alpha' + V)\, D^1] = Tr_1[hD^1] \qquad (28)$$

we may set up the following auxiliary functional in order to perform the second variation implicit in Eq. (20):

$$\Omega_v[D^1] = H_v[D^1] - \mu\left(\sum_{i=1}\tau_i^2 - n\right) - \sum_{i=1}\sum_{j=1}\beta_{ij}\left(<\psi_i|\psi_j> - \delta_{ij}\right) \tag{29}$$

The normalization condition on $D^1(1,1')$ is incorporated through the Lagrange multiplier μ and the orthonormality condition on the one-particle orbitals $\{\psi_i\}$, through the β_{ij}'s.

The operator $\alpha(1)$ depends upon the λ_i's and the orbitals $\psi_i(1)$. Hence, the auxiliary functional $\Omega_v[D^1]$ also depends upon these quantities. But instead of the λ_i's, it turns out that the natural variable ρ of the problem are the square roots of the occupation numbers ($\tau_i = \lambda_i^{1/2}$). This conclusion was obtained when the present method was applied to a 2-particle system, for which the exact coupled equations for the natural orbitals may be derived in a more direct way /16,17/.

Bearing the above in mind, we obtain the following sets of Euler-Lagrange equations

$$\frac{\delta\Omega_v[D^1]}{\delta\tau_i} = 0 \qquad \forall\ i \in \{1,\ldots,m\};\quad m > N \tag{30}$$

$$\frac{\delta\Omega_v[D^1]}{\delta\psi_i^*(1)} = 0 \qquad \forall\ i \in \{1,\ldots,m\};\quad m > N \tag{31}$$

from Eq.(28) we have

$$H_v[D^1] = Tr_1[hD^1] = \sum_{i=1}^{m} h_{ii}\left[\{\tau_i\},\{\psi_i\}\right]\tau_i^2 \tag{32}$$

thus, taking the derivative of Eq.(32) with respect to τ_K one gets

$$\mu_\tau \equiv \frac{\delta\mu_v[D^1]}{\delta\tau} = \sum\frac{\delta h_{ij}}{\delta\tau}\tau_i^2 = 2h_{KK}\tau_K \tag{33}$$

using that Eq. (30) may be rewritten in matrix form as

$$\begin{bmatrix} (h_{11}-\mu) & R_{12} & \cdots & R_{1m} \\ R_{21} & (h_{22}-\mu) & \cdots & R_{2m} \\ \vdots & \vdots & & \\ R_{m1} & R_{m2} & & (h_{mm}-\mu) \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_m \end{bmatrix} = 0 \qquad (34)$$

where

$$R_{ij} = \frac{\delta h_{ii}[\{\tau_j\},\{\psi_j\}]}{\delta \tau_j}\ \tau_i \qquad (35)$$

Eq. (34) is an eigenvalue equation which leads to a polynomial in μ for each $\mu^{(p)}$, there corresponds the column vector $\vec{\tau}^{(p)}$.

The orbital equations are obtained by explicitly varying Eq. (31). This variation yields

$$\mathrm{Tr}\left(\frac{\delta h^{(p)}}{\delta \psi_K^*(1)}\, D^{1(p)}\right) + \left(\tau^{(p)}\right)^2 \int d1\, h^{(p)}(1')\, \psi_K^{(p)}(1')\Big|_{1'\to 1} - \sum_{i=1}^{m} \beta_{iK}^{(p)} \psi_i^{(p)}(1) = 0 \qquad (36)$$

Notice that these orbital equations depend upon the particular "state" p described by $\vec{\tau}^{(p)}$. Furthermore, since Eqs. (34) and (36) must be solved self-consistently, it follows then that associated to each $\mu^{(p)}$ there must be a 1-matrix $D^{1(p)}$ for state p. See Section 5 below for further discussion on this point.

4. CONNECTION WITH THE KOHN-SHAM EQUATIONS.

Although at first sight the system of equations described by Eqs. (34) and (36) do not seem to resemble the usual Kohn-Sham orbital equations, the correspondence can be made evident by remembering that according to the first equality of Eq. (26). The universal functional $F[D^1]$ may be computed from $\mathrm{Tr}[\hat{H}_o D_o^N]$, where D_o^N is the N-matrix at the extreme point of variation. Thus, we have

$$\mathrm{Tr}[\hat{H}_o D_o^N] = \int d1 \left(-\frac{1}{2}\nabla_{1'}^2\right) D_o^1(1,1')\big|_{1'\to 1} + \int d1 \int d2 \frac{D_o^2(1,2;1,2)}{|\vec{r}_1 - \vec{r}_2|} \tag{37}$$

Using Eq. (37) we can put

$$H_v[D^1] = \int d1 \left(-\frac{1}{2}\nabla_{1'}^2 + V(\vec{r}_1')\right) D_o^1(1,1')\big|_{1'\to 1} + \int d1 \int d2 \frac{D_o^2(1,2;1,2)}{|\vec{r}_1 - \vec{r}_2|} \tag{38}$$

and compute

$$\frac{\delta H_v[D^1]}{\delta \psi_k^*(1)} = \lambda_k \int d1 \delta\psi_k^*(1) \left(-\frac{1}{2}\nabla_1^2 + V(\vec{r}_1)\right) \psi_k(1) + \int d1 \int d2 \frac{\delta D_0^2(1,2;1,2)}{\delta \rho(1)} \frac{\delta D_o^1(1,1)}{\delta \psi_k^*(1)} \tag{39}$$

where in the last term we have used the chain rule for the functional derivate plus the fact that $\rho_0(1) = D_o^1(1,1)$. It has been shown by Fritsche /18/ that the second term in Eq. (39) may be rewritten in terms of the effective potential

$$U(1) \equiv \int d2 \frac{\delta D^2(1,2;1,2)}{\delta \rho(1)} \tag{40}$$

Using Eqs. (39) and (40) we may transform Eq. (31) (or Eq. (30)) into

$$\lambda_k^{(p)} \left(-\frac{1}{2}\nabla_1^2 + V(\vec{r}_1) + U^{(p)}(1)\right) \psi_k^{(p)}(1) = \sum_{i=1}^{m} \beta_{ik}^{(p)} \psi_i^{(p)} \tag{41}$$

Fixing the occupation numbers (assuming that p = 0 describes the ground state) as follows

$$\lambda_k^{(o)} = 1 \qquad k = 1, \ldots, N$$
$$\lambda_k^{(o)} = 0 \qquad k > N \tag{42}$$

we set the following orbital equation

$$\left(-\frac{1}{2}\nabla_1^2 + V(\vec{r}_1) + U^{(o)}(1)\right)\psi_k^{(o)}(1) = \sum_{i=1}\beta_{ik}^{(o)}\psi_i^{(o)}(1) \quad (43)$$

Diagonalizing the $\underline{\beta}^{(o)}$ matrix by means of the transformations

$$\vec{\phi} = \underline{U}\,\vec{\psi} \quad , \quad \underline{\varepsilon} = \underline{U}\,\underline{\beta}\,\underline{U}^{+} \quad (44)$$

one finally obtains the Kohn-Sham equations

$$\left(-\frac{1}{2}\nabla_1^2 + V(\vec{r}_1) + U_{\phi}^{(o)}(1)\right)\phi_k(1) = \varepsilon_k^{(o)}\phi_k(1) \quad (45)$$

Notice however, that since the occupation numbers were fixed through Eq. (42), there is no companion equation, such as Eq. (34) to the usual Kohn-Sham Eq. (45). In this sense, Eq. (45) only refers to the ground state /19/.

5. MULTICONFIGURATION, STATIONARY STATES, CALCULATIONS AND THIRD THEOREM IN DENSITY FUNCTIONAL THEORY.

From the considerations mentioned above where we discussed that even if the name *density* functional theory suggests the use of the electronic density itself and alone, when one comes to the calculation of all relevant properties, the most successful practical schemes (eg. Kohn & Sham), one way or another, rely on the use of the first order density matrix $D^1(1,1')$ as mentioned above. Therefore, if an approximation to D^1 is computed anyhow, and having abandoned the conceptual advantage of using just the density ρ, one is tempted to extract full profit of the first order density matrix. (at a Hartree-Fock level the results can be obtained exactly). One practical calculational scheme is sketched here.

The first order density matrix has the multiconfigurational form

$$D^1(1,1') = \sum_A\sum_B C^{A*}C^B \sum_{i\in A}\sum_{j\in B}\Delta_{AB}^{ij}\phi_i^*(1)\phi_j(1') \quad (46)$$

where $\{C^A\}$ is the set of multiconfigurational parameters

of orthonormal wave functions out of which the configurations A are constructed:

$$\Delta_{AB}^{ij} = \begin{cases} 1 \text{ if } \{i' \text{ in A}\} - i = \{j' \text{ in B}\} - j \\ 0 \text{ otherwise} \end{cases} \tag{47}$$

(for bosons the values of Δ_{AB}^{ij} depends on the occurrence numbers of the ϕ_i in A).

The minimization procedure with respect to the ϕ_i^* yields

$$k_i^f + V_{eff,i}^f - \mu\, a_i^f = 0 \tag{48}$$

where

$$V_{eff,i}^f = \frac{\partial \iint V_{eff} D^1(1,1')dr_1 dr_{1'}}{\partial \phi_i^*} \tag{49}$$

and

$$k_i^d = \frac{\partial \left(-\frac{1}{2} \int \nabla_1^2 D^1(1,1')\big|_{1' \to 1} dr_1 \right)}{\partial \phi_i^*} \tag{50}$$

and

$$a_i^f = \sum_A \sum_B C^{A*} C^B \sum_{j \in B} \Delta_{AB}^{ij} \phi_j(1) \tag{51}$$

As a result we have that as the "kinetic energy" term is

$$-\sum_{AB} \sum_{i \in A} \sum_{j \in B} \left(C^{A*} C^B \Delta_{AB}^{ij} \phi_i^*(1) \nabla^2 \phi_j(1) + \text{h.c.} \right)/4 \tag{52}$$

And its contribution to the one-particle equation, that is, its derivative with respect to ϕ_i^*, will contain terms like

$$-C^{A*} C^{*A'} \Delta_{AA'}^{ij} \nabla^2 \phi_j(1) \tag{53}$$

not necessarily positive, depending on the relative signs of the C^A. Then the differential equations for the basis functions ϕ_i will be coupled. This is most important for molecules and solids where a LCAO scheme for the ϕ_i can be used. This coupling is an effect arising from the use of the full first order density matrix, where contributions to correlation will appear. If

natural orbitals are used for D^1 then (53) becomes diagonal, the projection of the resulting orbitals being used to weight the different configurations.

A second, and also very important consequence, is the appearence of local density fluctuations given by (46) when the non-diagonal $C^{A*}C^{B}\Delta^{ij}_{AB} \neq 0$. This point has been discussed by Fritsche. The solution of the variational equation (48) is to be found in conjuction to the ones given by $\partial\Omega/\partial C^{A} = 0$ where the coefficients C^{A} are optimized. In practice we can compute the total energy $E_B(\{C^A\}_B)$ and find a set of numerical minima for E_B each solution E_B corresponding to a stationary state of the system described by $\{C^A\}_B$.

Then it is clear that the internal symmetry of the system, through a set of parameters $\{C^A\}_B$ is a *necessary* part of the theory. The parameters C^A should be included (hidden or explicit) either in the construction of the density function (as in the Kohn-Sham procedure as a particular case) or in the energy functional itself.

We have called this statement the "Third Theorem" of density functional theory /19,20,21,22/. We have also discussed the effects of this internal structure of the density on the excitations of the system.

6. EXAMPLE: REAL AND CONFIGURATION SPACE WAVE FUNCTIONS OF f ELECTRONS IN CONDENSED MATTER.

6.1. Real Space Wave Functions.

For the study of rare earth and actinide elements containing systems we can start by considering again that the total wave function of a material which can be written as a fixed configuration interaction (CI) sum, /23,24/, where as usual each configuration is represented by one determinant constructed from appropriate single particle wave functions of different angular momenta. Each single particle wave function Φ_i is, on the other hand, constructed (usually) as a linear combination of atomic

The calculation of the single particle 4f-electron wave functions in atomic problems shows that even if these states are occupied only after the 5s, 5p and 6s orbitals, the charge distribution of the 4f is such that most of it is inside the sphere of maximum charge density of the 5s and 5p. This is also the case for the 5f electrons, in the actinides, relative to the 6s and 6p states. The physical reason for this behaviour is the dominant role played by the effective f potential $V_{\ell=3}(r)$, a radial potential well confining the f-wave function (for n f electrons) to a small region of space. The resultant f-wave function ψ_f can be appropriately called atomic-like. When the rare earth or actinide atom is in condensed matter, the boundary conditions and the potential well changes drastically, mainly in the outermost part of the Wigner-Seitz cell, through the superposition of the atomic potential wells, with a lower potential in the interstitial region. The change in the potential makes it necessary to include an extra contribution $\psi_{f'}$, , with large amplitude in the bonding region, to construct an appropriate f wave function $\tilde{\psi_f}$ in this material as:

$$\tilde{\psi_f} = \left(a_n \psi_f + \sqrt{1 - a_n^2}\, \psi_{f'}\right); \int |\psi_f|^2 d\tau = 1, \int |\psi_{f'}|^2 d\tau = 1 \qquad (54)$$

The auxiliary $\psi_{f'}$ wave function is not of atomic character, on the contrary it must have contributions outside the sphere where the 5s, 5p and 5d (or 6s, 6p and 6d for actinides) have their maxima. The $\psi_{f'}$ is (usually) taken to be orthogonal to ψ_f. The angular part is of course the same as the 4f (or 5f); in molecular and solid state calculations the $\psi_{f'}$ are usually termed "polarization" wave functions. Because the two components ψ_f and $\psi_{f'}$ are orthogonal, the f charge density will have two main contributions: the atomic-like part a_n^2 and the delocalized part $(1 - a_n^2)$.

The coefficient a_n is given the subscript n because it strongly depends on the total amount of f-character for a given

configuration f^n being considered. The lower the average energy of the f-electron band in the f^n configuration for that atom in condensed matter, the larger the a_n is expected to be ($a_n^2 \sim 1$). We can also think of ψ_f and $\psi_{f'}$ as "configurations", with the same amount of $\ell = 3$ character, which are mixed by the condensed matter boundary conditions.

6.2 Configuration Space Wave Functions.

Here we want to consider that the amount of f character in the rare earth or actinide can change either by pressure, alloying, excitations, etc. and then the total wave function Ψ could now be written in terms of $\underset{\sim}{\psi}^n_f$ as

$$\Psi = \sum_n C^n \underset{\sim}{\psi}^n_f \qquad (55)$$

Then in the procedure described in Section 5 we would have to optimize both the C^n and the a_n (perhaps written in the form C^{on} and C^{1n}). The sum $\sum_n |C^n|^2$ will then give the total amount of f character (better said $\ell = 3$ character) and the sum $\sum_n |C^n|^2 a_n^2$ the total amount of atomic-like f character. These numbers have been discussed by several authors (see references in /23,24/) and is related to mixed valence, valence fluctuations and the high electron correlation presented by these materials. The different conditions favor the contraction or expansion of the $\ell = 3$ auxiliary wave functions, arising from the extra terms in (53), and indirect changes in all angular momenta components because the f charge localizes or delocalizes.

ACKNOWLDEGEMENTS.

This work was partially supported by CONACYT, México, project Estructura Electrónica de Materiales de Alta Tecnología and by the Sistema Nacional de Investigadores, México. The technical assistance of Mrs. Irma Aragón is gratefully acknowldeged.

REFERENCES.

1. a) R.G. Parr, (1980). In "Horizons in Quantum Chemistry", (K. Fukui and B. Pullman, eds.), p. 5. Reidel, Dordrecht.

b) A.S. Banzai, and B.M. Deb, (1981). Rev. Mod. Phys. 53, 95.

c) N.H. March (1981). In "Theoretical Chemistry", A Specialist Periodical Report, (C. Thompson, ed.), 4, p. 92 Royal Society of Chemistry, London.

d) R.G. Parr, (1982) In "Electron Distributions and the Chemical Bond" , (P. Coppens and M.B. Hall, eds.), p. 95, Plenum, New York.

e) J. Keller and J.L. Gázquez (eds.) (1983) "Density Functional Theory" , Springer-Verlag, Berlin.

f) S. Lundqvist and N.H. March (eds.) (1983), "Theory of the Inhomogeneous Electron Gas", Plenum, New York.

g) J.P. Dahl and J. Avery (eds.) (1984) "Local Density Approximations in Quantum Chemistry and Solid State Physics" Plenum, New York.

h) R.M. Dreizler and J. da Providencia (eds.) (1985) "Density Functional Methods in Physics", Plenum, New York.

i) R.M. Erdhal and V.H. Smith (eds.) (1987) "Density Matrices and Density Functional", Reidel, Dordrecht.

2. For a synthesis given in terms of local-scaling transformations, see E.S. Kryachko, I. Zh. Petrov and M.V. Stoitsov (1987), Int. J. Quantum Chem. 32, 473 ; E.S. Kryachko and E.V. Ludeña (1987), Phys. Rev. A 35h, 957 ; E.S. Kryachko and E.V. Ludeña (1989) In "Condensed Matter Theory" 4, (J. Keller, ed.), Plenum, New York 1989.

3. A. J. Coleman, (1961), Can. Math. Bull., 4, 209

4. A.J. Coleman. In Ref. 1-i, p. 5

5. A.J. Coleman, (1963), Rev. Mod. Phys. 35, 668.

6. T.T. Nguyen-Dang, E.V. Ludeña and Y. Tal. (1985) , J. Mol. Struct. (Theochem) 120, 247.

7. M. Levy (1979), Proc. Natl. Acad. Sci. (U.S.A.) 76, 6062.

8. E.H. Lieb (1983), Int. J. Quantum Chem. 24, 243.
9. P. Hohenberg and W. Kohn (1964), Phys. Rev. 136B, 864.
10. W. Kohn and L.J. Sham (1965), Phys. Rev. A140, 1133.
11. H. Englisch and R. Englisch (1984), Phys. Stat. Sol. (b) 123, 711; Physica 121A, 253 (1983). J.T. Chayes, L. Chayes and M.B. Ruskai (1985), J. Stat. Phys. 38, 497.
12. N. Nakatsuji and R.G. Parr (1975), J. Chem. Phys. 63, 1112.
13. E.V. Ludeña and A. Sierraalta (1985), Phys. Rev. A32, 19.
14. E.V. Ludeña, (1987), In "Density Matrices and Density Functionals", (R. Erdahl and V.H. Smith, eds.), Reidel, Dordrecht, p. 289.
15. P.O. Lowdin (1955), Phys. Rev. 97, 1474.
16. E.R. Davidson (1976). In "Reduced Density Matrices in Quantum Chemistry", Academic Press, New York, p. 73.
17. E.V. Ludeña (1985), J. Mol. Struct. (Theochem) 123, 371.
18. L. Fritsche (1985), Phys. Rev. B33, 3976
19. J. Keller, C. Amador and C. de Teresa (1986), In "Condensed Matter Theories" (B. Malik, ed.), Plenum, New York, Vol. 1, 195; J. Keller, C. Amador, C. de Teresa and J.A. Flores (1986). In "Condensed Matter Theories" (P. Vashishta, ed.), Plenum, New York, Vol. 2, p.131.
20. J. Keller (1986), Int. J. Quantum Chem., Quantum Chem. Symp. 20, 767; J. Keller and E. Ludeña (1987), Int. J. Quantum Chem., Quantum Chem. Symp. 21, p.171.
21. J. Keller and C. Amador (1987). In "Condensed Matter Theories" (J. Arponen, ed.), Plenum, New York, Vol. 3, p.201.
22. J. Keller (1988), "Journal of Molecular Structure" (Theochem) Elsevier Science Pubs., Vol. 166, 51.
23. J. Keller, C. de Teresa, J. Schoenes (1985), Solid State Comm. 56, 871.
24. J. Keller and C. de Teresa (1988). In "The Challenge of d and f Electrons" (D.R. Salahub and M. Zerner, eds.), American Chemical Society

CONSTRAINED-SEARCH FORMULATION AND RECENT COORDINATE SCALING IN DENSITY-FUNCTIONAL THEORY

MEL LEVY
Chemistry Department and Quantum Theory Group
Tulane University
New Orleans, Louisiana 70118
U.S.A.

Abstract

The constrained-search formulation, which simplifies and generalizes the original Hohenberg-Kohn theorem and solves the v-representability problem, is briefly reviewed. The constrained-search formulation is then combined with coordinate scaling to derive bounds for the exact exchange-correlation functional. The asymptotic bounds are not generally satisfied by commonly employed approximations. Coordinate scaling is then employed to present formulas which extract ground-state energies from experimental electron densities and to present a set of three necessary and sufficient conditions for the exact orbital-generated Kohn-Sham exchange potential which is a local operator. Finally, coordinate scaling equality requirements are revealed for the correlation hole.

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

TABLE OF CONTENTS

1. INTRODUCTION

Hohenberg and Kohn[1] ushered in modern density functional theory by asking and answering the following question: what is it that Thomas-Fermi[2] and extended Thomas-Fermi functionals are trying to approximate? It seems suprising that it took almost forty years for this question to be asked. In part because the question was asked and answered, and in part because of its numerical success, density functional theory has blossomed during the twenty five years since the Hohenberg-Kohn theorems. Witness the large number of density functional papers at the Sanibel Symposia and at the March Meetings of the American Physical Society. Density functional sessions now constitute regular parts of the advertised programs of these annual conferences.

The present chapter first briefly reviews the simplified and generalized "constrained-search" proofs[3-5] of the Hohenberg-Kohn theorems for ground states. This chapter then combines the constrained-search orientation with coordinate scaling to derive various properties of the exact exchange-correlation functional, E_{xc}, for ground-state calculations. Several of these derived properties are not satisfied by commonly employed approximations to E_{xc}. This fact should be considered when one attempts to improve upon the existing exchange-correlation functionals. Further implementation of coordinate scaling in this chapter results in formulas which generate energies from experimental electron densities and results in a theorem for the exact orbital-generated Kohn-Sham exchange potential. The chapter concludes by revealing coordinate scaling equality requirements for the correlation hole.

2. SIMPLIFIED AND GENERALIZED PROOFS OF HOHENBERG-KOHN THEOREMS A AND B FOR GROUND STATES.

2.1. Constrained-Search Proof of Theorem A. Ground-State Density Determines "Everything."

Consider N interacting electrons in a local spin-independent external potential v. The Hamiltonian is

$$H + T + V_{ee} + \sum_i v(\vec{r}_i) \tag{1}$$

where T and V_{ee} are, respectively, the kinetic and electron-electron repulsion operators. According to Hohenberg-Kohn[1] Theorem A, a given ground-state density, n_o, determines H within an additive constant and thus determines all the ground-state and excited-state properties of H.

By means of the "constrained-search" formulation,[3-5] the proof of Theorem A becomes quite direct. A proof by construction is employed. That is, the proof actually shows how one determines, in principle, H from n_o.

Note first that all those antisymmetric wavefunctions, Ψ, which yield the given n_o possess the same expectation value with respect to any local-multiplicative external potential v. (The electron-nuclear attraction operator is an example of a local external potential.) Consequently, of all those Ψ's which yield n_o, the ground-state Ψ is obtained as the one which minimizes $\langle T + V_{ee} \rangle$. (Call this wavefunction $\Psi^{min}_{n_o}$.) If degeneracies exist such that more than one Ψ gives this minimum, then all the ground-state Ψ's are obtained, one at a time, by the above procedure. Furthermore, all the Ψ's which give the constrained minimum must obviously be ground states of the same local external potential.

Once a ground-state, $\Psi^{min}_{n_o}$, has been determined, then the total local-multiplicative potential operator, $V = V_{ee} +$

$\sum_i v(\vec{r}_i)$, is obtained within an additive constant by inverting $H\Psi^{min}_{n_o} = E_o\Psi^{min}_{n_o}$. That is

$$V\text{-}E_o = -T\Psi^{min}_{n_o}/\Psi^{min}_{n_o} \tag{2}$$

where the operator T is specified completely by the number of electrons which, in turn, is determined from n_o by integration. Hence, H has been determined, within an additive constant, and the proof is complete. (Observe, incidentally, that since H is specified only to within an additive constant, E_o cannot be determined without further information. However, further information is not required for excitation energies which involve only energy differences.)

2.1.2. Hartree-Fock Density Determines "Everything".

Hohenberg-Kohn Theorem A has been extended to the Hartree-Fock density,[6-8] n_{HF}, in a rigorous manner[8] through the Fock Hamiltonian.[8] (See also Ref. 3). Consequently, n_{HF} contains all the information about H, including the correlation energy.[7,8] That is,

$$n_{HF} \rightarrow v \rightarrow (E_o\text{-}E_{HF}) = \tilde{E}_c[n_{HF}]. \tag{3}$$

The existence of the universal correlation energy functional, $\tilde{E}_c[n_{HF}]$, provides hope for a practical addition of $\tilde{E}_c[n_{HF}]$ to E_{HF} after one has performed a traditional Hartree-Fock calculation for E_{HF} and n_{HF}; one does not need the exact n_o! Namely,[7,8]

$$E_o = E_{HF} + \tilde{E}_c[n_{HF}]. \tag{4}$$

However, Hartree-Fock calculations are impractical for very large molecules and solids so that in these cases it is perhaps best, at present, to employ traditional density-functional techniques. This brings us to the variational Hohenberg-Kohn Theorem.[1]

2.2. Constrained-Search Proof of Theorem B. Existence of Universal Variational Functional for $\langle T + V_{ee} \rangle$.

Hohenberg-Kohn Theorem B states that there exists a universal F[n] such that

$$E_o = \min_n \left(\int v(\vec{r}) n(\vec{r}) d^3r + F[n] \right). \tag{5}$$

F[n] is universal in that its value is independent of the particular v of interest. Again, the constrained-search formulation provides a direct proof. Simply define F[n] to be[3-5]

$$F[n] = \min_{\Psi \to n} \langle \Psi | T + V_{ee} | \Psi \rangle \tag{6}$$

or

$$F[n] = \langle \Psi_n^{min} | T + V_{ee} | \Psi_n^{min} \rangle, \tag{7}$$

where Ψ_n^{min} minimizes $\langle T + V_{ee} \rangle$ and is constrained to yield the trial n. Equation (5) follows directly because Eq. (6) implies that

$$\min_n \left(\int v(\vec{r}) n(\vec{r}) d^3r + F[n] \right) \tag{8}$$

$$= \min_n \left(\min_{\Psi \to n} \langle \Psi | H | \Psi \rangle \right)$$

$$= \min_{\Psi} \langle \Psi | H | \Psi \rangle = E_o$$

It also follows from the constrained-search proof[3] of Theorem A that

$$E_o = \int v(\vec{r}) n_o(\vec{r}) d^3r + F[n_o]. \tag{9}$$

Harriman[9] has shown that N-representability is no problem for any trial n and Lieb[10] has shown that Ψ_n^{min} always exists. In addition, Valone[11] has provided an ensemble-search generalization of Eqs. (6) and (7). Moreover, Perdew, Parr, Levy, and Balduz, Jr. have provided a fractional electron number generalization in connection with their discovery of the "derivative discontinuity" of the energy.[12] The definition of F[n] in equations (6) and (7) and its ensemble-search generalization follow in the spirit of Percus' definition[13] of a universal kinetic energy functional for independent fermion systems. Percus, however, did not address the v-representability problem. The F[n] in Eqs. (6) and (7) solves the v-representability problem by liberating the original Hohenberg-Kohn functional from the v-representability restriction.[3-5]

3. PARTITIONING OF F[n].

For the purpose of achieving an accurate approximation to F[n] for computational uses, it is convenient to partition F[n] according to[14]

$$F[n] = T_s[n] + U[n] + E_{xc}[n] \tag{10}$$

where

$$T_s[n] = \langle \Phi_n^{min} | T | \Phi_n^{min} \rangle, \tag{11}$$

where U[n] is the classical electron-electron repulsion

energy, where Φ_n^{min} is that wavefunction which yields n and minimizes just $\langle T \rangle$, and where $E_{xc}=E_x + E_c$ is the exchange-correlation functional.

The exchange component, E_x, is normally defined in density-functional theory as[15]

$$E_x[n] = \langle \Phi_n^{min} | V_{ee} | \Phi_n^{min} \rangle - U[n] \tag{12}$$

which gives the correlation component as[15]

$$E_c[n] = \langle \Psi_n^{min} | T + V_{ee} | \Psi_n^{min} \rangle - \langle \Phi_n^{min} | T + V_{ee} | \Phi_n^{min} \rangle. \tag{13}$$

There are other ways to partition E_{xc}. This particular partitioning, however, will prove useful here for the derivations of coordinate scaling properties for the whole E_{xc}. Also, see Sabin and Trickey[16] for a detailed analysis of correlation contributions in local density models.

The wavefunction Φ_n^{min} is quite often a single determinant and a ground state of a non-interacting system, especially when n is the interacting ground-state density of a closed-shell atom or molecule. In this common case Φ_n^{min} is the same as the original Kohn-Sham non-interacting single determinant ground-state wavefunction. In any case, there are slightly different ways of defining $T_s[n]$, but virtually all the definitions involve a constrained-search minimization with just $\langle T \rangle$.

In most practical calculations E_x and E_c are approximated but, at present, T_s is computed exactly by the use of orbitals in the Kohn-Sham procedure.[14]

4. BOUND FOR DENSITY-FUNCTIONAL EXCHANGE.

Most often approximations to the E_x of Eq. (12) are compared to the traditional exchange as defined in

Hartree-Fock theory. The later exchange, which shall be called E_x^{HF}, is defined by

$$E_x^{HF}[n] = \langle\Phi_n^{HF}|V_{ee}|\Phi_n^{HF}\rangle - U[n] \tag{14}$$

where Φ_n^{HF} is that single determinant which yields n and minimizes[3,6] $\langle T + V_{ee}\rangle$. Exact values for E_x^{HF} are known for many more cases than with E_x. This is the main reason why the value of E_x^{HF} is so often employed to judge the quality of an approximate E_x.

It is usually tacitly assumed in the literature that $E_x[n]$ approximately equals $E_x^{HF}[n]$, which is probably generally close to being true. It is nevertheless worthwhile to learn more about the relationship between E_x and E_x^{HF}, especially since approximations to E_x are getting better all the time. With this in mind, it is now asserted that[17]

$$E_x[n] \geq E_x^{HF}[n] \tag{15}$$

when Φ_n^{min} is a single determinant.

To prove[17] Eq. (15), use the fact that the definitions of Φ_n^{min} and Φ_n^{HF} dictate that

$$\langle\Phi_n^{min}|T + V_{ee}|\Phi_n^{min}\rangle \geq \langle\Phi_n^{HF}|T + V_{ee}|\Phi_n^{HF}\rangle \tag{16}$$

when Φ_n^{min} is a single determinant and that always

$$\langle\Phi_n^{HF}|T|\Phi_n^{HF}\rangle \geq \langle\Phi_n^{min}|T|\Phi_n^{min}\rangle. \tag{17}$$

The combination of Eqs. (12), (14), (16), and (17) gives Eq. (15).

Equation (15) can be expressed equivalently as[17]

$$\lim_{\gamma\to\infty} \gamma^{-1} E_{xc}[n_\gamma] \geq E_x^{HF}[n] \tag{18}$$

where $n_\gamma(\vec{r}) = \gamma^3 n(\gamma\vec{r})$, because the left-hand-side of Eq. (18) equals E_x as proven later [see Eq. (40)]. Equation (18) is useful when the approximate E_x is not clearly discernible within the approximate E_{xc}. In any case, it is noteworthy that approximate functionals do not often satisfy Eq. (15) or Eq.(18).

5. COORDINATE SCALING THEOREM.

Coordinate scaling has proven to be quite helpful for understanding the forms of $T_s[n]$, $E_x[n]$, and $E_c[n]$, and for approximating these functionals. In the development below use shall be made of the following scaling theorem of Levy and Perdew.[18] If $\Psi_n^{\min,\alpha}$ is that antisymmetric wavefunction which yields n and minimizes $\langle T + \alpha V_{ee} \rangle$, then $\beta^{3N/2}\Psi_n^{\min,\alpha}(\beta\vec{r}_1, \beta\vec{r}_2, \ldots \beta\vec{r}_N)$ minimizes $\langle T + \beta\alpha V_{ee} \rangle$ and yields the scaled density n_β, where $n_\beta(\vec{r}) = \beta^3 n(\beta\vec{r})$.

5.1. T_s and E_x Scale Homogeneously.

Observe that $\Phi_n^{\min} = \Psi_n^{\min,o}$. Since $\Phi_n^{\min}$ corresponds to $\alpha = 0$, it follows from the above theorem[18] that $\lambda^{3N/2}\Phi_n^{\min}(\lambda\vec{r}_1, \ldots, \lambda\vec{r}_N)$ is that wavefunction which yields n_λ and minimizes just $\langle T \rangle$. In other words,

$$\Phi_{n_\lambda}^{\min}(\vec{r}_1, \ldots, \vec{r}_N) = \lambda^{3N/2}\Phi_n^{\min}(\lambda\vec{r}_1, \ldots, \lambda\vec{r}_N). \tag{19}$$

This means that[18]

$$T_s[n_\lambda] = \lambda^2 T_s[n] \tag{20}$$

and[18]

$$E_x[n_\lambda] = \lambda E_x[n] \tag{21}$$

where it has been taken into consideration that T is homogeneous of degree -2 and V_{ee} is homogeneous of degree -1. (Unlike T and E, the true kinetic energy and the true electron-electron repulsion energy do not[18] scale homogeneously. It is thus pertinent that Levy, Yang, and Parr[19] have defined a modified true kinetic energy functional and a modified true electron-electron repulsion functional which exhibit homogeneous scaling.)

According to Eq. (20) the local density component of T_s must have the form $\int n^{5/3}$ (Thomas-Fermi form), and according to Eq. (21) the local density component of E_x must have the form $\int n^{4/3}$ (Dirac form). The forms of nonlocal improvements upon $\int n^{5/3}$ and $\int n^{4/3}$ are also dictated, to a significant extent, by Eqs. (20) and (21). For instance[20]

$$\tilde{T}_s[n] = a(N)\int n^{5/3} + b(N)\int n^{-1}|\vec{\nabla}n|^2 d^3r + c(N)\int n(\vec{r})\vec{r}\cdot\vec{\nabla}[n^{-2/3}|\nabla n|]d^3r \tag{22}$$

where the second term is von-Weizsäcker-like and the third term helps produce a correct property of the functional derivative (it has been shown[21,22] that $\frac{\delta T_s[n]}{\delta n(\vec{r})} \geq -\frac{1}{2}\nabla^2 n^{1/2}/n^{1/2}$). Each term in Eq. (22) scales as λ^2. (The tilde is used to signify that the functional is an approximate one.) Attractive nonlocal exchange functionals now include the one by Perdew and Wang:[23]

$$\tilde{E}_x[n] = -\beta\int d^3r n^{4/3}(\vec{r})\, Q(s), \tag{23}$$

where $Q(s) = (1 + as^2 + bs^4 + cs^6)^{1/15}$, and where $s = |\vec{\nabla}n|/n^{4/3}$. The $\tilde{E}_x$ in Eq. (23) scales correctly. Namely,

$$\tilde{E}_x[n_\lambda] = -\beta\lambda\int d^3(\lambda r) n(\lambda\vec{r})^{4/3} Q(s_\lambda) = \lambda E_x[n] \tag{24}$$

because

$$s_\lambda = |\nabla n_\lambda|/n_\lambda^{4/3} = |\nabla(\lambda\vec{r})n(\lambda\vec{r})|/n(\lambda\vec{r})^{4/3}. \tag{25}$$

It is important to note that by the Coordinate Scaling Theorem[18] $E_x^{HF}[n]$ does <u>not</u> satisfy Eq. (21) because Φ_n^{HF} minimizes $\langle T + V_{ee}\rangle$ and not just $\langle T\rangle$. Hence, if one's approximate exchange functional satisfies Eq. (21) it follows that one is really approximating E_x and not E_x^{HF}. This point is not always appreciated.

5.2. E_c Does Not Scale Homogeneously.

Unlike T_s and E_x, E_c does <u>not</u> exhibit homogeneous scaling.[18] The reason for this is the fact that

$$\Psi_{n_\lambda}^{min} \neq \lambda^{3N/2}\Psi_n^{min}(\lambda\vec{r}_1, \ldots, \lambda\vec{r}_N) \tag{26}$$

because, by the Coordinate Scaling Theorem,[18] the scaled wavefunction $\lambda^{3N/2}\Psi_n^{min}(\lambda\vec{r}_1, \ldots, \lambda\vec{r}_N)$ minimizes $\langle T + \lambda V_{ee}\rangle$ and <u>not</u> $\langle T + V_{ee}\rangle$. As a result,

$$F[n_\lambda] < \lambda^2 T[n] + \lambda V_{ee}[n];\ \lambda \neq 1, \tag{27}$$

where $T[n] = \langle\Psi_n^{min}|T|\Psi_n^{min}\rangle$ and $V_{ee}[n] = \langle\Psi_n^{min}|V_{ee}|\Psi_n^{min}\rangle$. Equation (27) leads to the following inequalities:[18]

$$\begin{aligned} &E_c[n_\lambda] > \lambda E_c[n];\ \lambda > 1 \\ &E_c[n_\lambda] < \lambda E_c[n];\ \lambda < 1 \end{aligned} \tag{28}$$

$$\lambda^{-1}E_c[n_\lambda] \leq \lambda\{(\partial E_c[n_\gamma]/\partial\gamma)_{\gamma=1} - E_c[n]\} + 2E_c[n] - (\partial E_c[n_\gamma]/\partial\gamma)_{\gamma=1} \quad (29)$$

Also, from $V_{ee}[n_\lambda] \geq 0$ and from the fact that $T_s[n] \leq T[n]$, it follows that

$$\lambda^{-1}E_c[n_\lambda] \geq -U[n]-E_x[n]. \quad (30)$$

Equations (29) and (30) dictate limiting bounds[24] on E_c:

$$2E_c[n] - (\partial E_c[n_\gamma]/\partial\gamma)_{\gamma=1} \geq \lim_{\lambda\to 0} \lambda^{-1}E_c[n_\lambda] \geq - U[n] - E_x[n] \quad (31)$$

As a tribute to the local density approximation (LDA), it should be stressed that the LDA has satisfied Eqs. (28-31) for all n's tested thus far.[18]

5.3. Asymptotic Scaling Bound For E_c Not Satisfied By LDA.

Quite recently asymptotic scaling requirements for E_c have been derived from the observation that

$$E_c[n_\lambda] = \langle\Psi^{\min}_{n_\lambda}|T + V_{ee}|\Psi^{\min}_{n_\lambda}\rangle - \langle\Phi^{\min}_{n_\lambda}|T + V_{ee}|\Phi^{\min}_{n_\lambda}\rangle \quad (32)$$

implies[24]

$$E_c[n_\lambda] = \lambda^2[\langle\Psi^{\min,\alpha}_n|T|\Psi^{\min,\alpha}_n\rangle - \langle\Phi^{\min}_n|T|\Phi^{\min}_n\rangle] + \lambda[\langle\Psi^{\min,\alpha}_n|V_{ee}|\Psi^{\min,\alpha}_n\rangle - \langle\Phi^{\min}_n|V_{ee}|\Phi^{\min}_n\rangle] \quad (33)$$

because[24]

$$\Psi^{\min}_{n_\lambda} = \lambda^{3N/2}\Psi^{\min,\alpha}_n(\lambda\vec{r}_1, \ldots, \lambda\vec{r}_N), \quad (34)$$

where $\alpha = \lambda^{-1}$. Focus upon very small $|\alpha|$ and first consider when $\Psi_n^{\min,\alpha}$ may be expanded in a power series in α:

$$\Psi_n^{\min,\alpha} = \Phi_n^{\min} + \sum_{m=1}^{\infty} \alpha^m g_m[n]. \tag{35}$$

Substitute Eq. (35) into Eq. (33) to yield

$$E_c[n_\lambda] = a[n] + b[n]\ \lambda^{-1} + c[n]\ \lambda^{-2} + \ldots \tag{36}$$

which dictates that $E_c[n_\lambda]$ is bounded as $\lambda \to \infty$.

For a common and important situation we have knowledge about the bound on $\lim_{\lambda\to\infty} E_c[n_\lambda]$. Consider

$$H_s^\alpha = T + \sum_{i=1}^{N} v_s(\vec{r}_i) + \alpha\, V_{ee} \tag{37}$$

Call E_s^α the ground-state energy of H_s^α. Now focus upon the $\alpha \to o$ limit and express E_s^α in the form of the following perturbation expansion:

$$E_s^\alpha = E_o + \alpha E_1 + \alpha^2 E_2 + \alpha^3 E_3 + \ldots \tag{38}$$

It has recently been shown that[25]

$$\lim_{\lambda\to\infty} E_c[n_\lambda] \geq E_2 \tag{39}$$

when $\Phi_n^{\min}$ is the non-degenerate ground-state of H_s^o. The constant E_2 is already known for densities composed of hydrogenic orbitals. For instance, $E_2 = -0.16$ a.u. when n is the two-electron ground-state density of $H_s^o = T - \sum_{i=1}^{2} r_i^{-1}$.

The LDA does not satisfy Eq. (36) or Eq. (39) because the LDA expression for E_c is unbounded as $\lambda \to \infty$. This fact should be considered when one is trying to improve upon the LDA. (The result that the LDA goes as -log λ when $\lambda \to \infty$ arises from the Gell-Mann-Brueckner high-density formula.)

5.4. Projection of E_x and E_c from E_{xc}.

Not all approximate exchange-correlation functionals are such that an exchange component can be easily separated from a correlation component. For the purpose of testing how an approximate E_{xc} scales it may thus be necessary to project the corresponding exchange and correlation components from the approximate E_{xc}. With this in mind, observe the Eqs. (36) and (39) reveal that $\lim_{\gamma\to\infty} \gamma^{-1}E_c[n_\gamma] = 0$. This implies[17,24]

$$E_x[n] = \lim_{\gamma\to\infty}\gamma^{-1}E_{xc}[n_\gamma] \tag{40}$$

since $\gamma^{-1}E_x[n_\gamma] = E_x[n]$ for all γ. Accordingly,[17,24]

$$E_c[n] = E_{xc}[n] - \lim_{\gamma\to\infty} \gamma^{-1}E_{xc}[n_\gamma]. \tag{41}$$

With Eq. (41), Eq. (39) becomes[25]

$$\lim_{\lambda\to\infty} \{E_{xc}[n_\lambda] - \lim_{\gamma\to\infty} \gamma^{-1} E_{xc}[n_{\gamma\lambda}]\} \geq E_2 \tag{42}$$

where $n_{\gamma\lambda}(\vec{r}) = (\gamma\lambda)^3 n(\gamma\lambda\vec{r})$.

6. FORMULAS FOR GROUND-STATE ENERGIES FROM EXPERIMENTAL DENSITIES

Experimental electron densities are becoming ever more accurate by x-ray diffraction[26] and electron diffraction methods.[27] However, few total energies or cohesive energies of solids have been extracted from these densities. Consequently, it is the purpose of this section to reveal formulas that should be useful for such energy determinations.

The first expression that comes to mind, of course, is the one involving the whole F[n]. Namely

$$E_o^{total} = \int v_{en}(\vec{r}) n_o(\vec{r}) d^3r + F[n_o] + V_{NN} \tag{43}$$

where E_o^{total} is the total ground-state energy associated with ground-state density n_o for electron-nuclear attraction operator v_{en}. Also, E_o^{total} is the sum of the electronic energy, E_o, and the nuclear-nuclear repulsion energy, V_{NN}. Now, since F must be approximated, it is worthwhile to put forth formulas which do not involve the whole of F. Coordinate scaling shall now be invoked for this purpose.

Since n_o is a ground-state density it follows that

$$\frac{\partial}{\partial \lambda} \left(\int v_{en}(\vec{r}) n_o^{\lambda}(\vec{r}) d^3r + F[n_o^{\lambda}] \right)_{\lambda=1} = 0, \tag{44}$$

where $n_o^{\lambda}(\vec{r}) = \lambda^3 n_o(\lambda \vec{r})$. The implementation of Eq. (44) with consideration of the homogeneous scaling properties of T_s,

E_x, and U, lead to

$$2T_s[n_o] + \int v_{en}(\vec{r})\ [\partial n_o^{\lambda}(\vec{r})/\partial\lambda]_{\lambda=1} d^3r +$$

$$U[n_o] + E_x[n_o] + (\partial E_c[n_o^{\lambda}]/\partial\lambda)_{\lambda=1} = 0. \qquad (45)$$

Equation (45) allows us to completely eliminate either T_s or E_x from Eq. (43). If, at any equilibrium nuclear configuration, we eliminate T_s then Eq. (43) becomes

$$E_o^{total} = \tfrac{1}{2} \int v_{en}(\vec{r})\ n_o(\vec{r})\ d^3r + \tfrac{1}{2} U[n_o] +$$

$$\tfrac{1}{2} E_x[n_o] + \tfrac{1}{2} V_{NN} + E_c[n_o] - \tfrac{1}{2} (\partial E_c[n_o^{\lambda}]/\partial\lambda)_{\lambda=1}. \qquad (46)$$

On the other hand, at an equilibrium nuclear configuration, if we eliminate E_x then Eq. (43) becomes

$$E_o^{total} = - T_s[n_o] + E_c[n_o] - (\partial E_c[n_o^{\lambda}]/\partial\lambda)_{\lambda=1}. \qquad (47)$$

Numerical estimates on atoms with Eqs. (46) and (47) have been encouraging.[28]

Whether one decides to use Eqs. (43), (46), or (47) should depend upon the situation. If F[n] were known exactly, then it would perhaps be best to employ Eq. (43) for approximate ground-state densities in order to take advantage of the stationary characteristic of the energy functional. However, F[n] will never be known exactly for computational purposes. Hence, Eqs. (46) and (47) should be seriously considered as alternatives to Eq. (43). Presently, approximations to T_s appear to be significantly less accurate[29] than approximations to E_x, in part because T_s is usually much larger in magnitude than either E_x or E_c (In the

Ne atom, for example, $T_s \simeq$ -128 a.u., $E_x \simeq$ -12 a.u., and $E_c \simeq$ -0.4 a.u.). For this Eq. (46) is probably the best to use at the present time. In Eq. (46), not only has T_s been removed but also the error in E_x in Eq. (46) is <u>half</u> that in Eq. (43) when n_o is employed.[28]

In the future, it is certainly possible that T_s might be computed as accurately as needed by the explicit determination of $\Phi^{min}_{n_o}$. Promising techniques for this purpose have recently been put forth by Aryasetiawan and Stott,[30] Levy and Goldstein,[28] and by Cioslowski.[31] It should be noted that Cioslowski's procedure[31] constitutes an explicit realization of the constrained-search method. Observe also that Eq. (47) does not require knowledge of E_x at all! Hence, perhaps Eq. (47) is the one for the future.

7. E_x FROM v_x.

There has recently been a rebirth of interest in improvements upon the original orbital-generated Slater potential,[32] as shall be discussed in the next section. Use shall be made of the following relation of Levy and Perdew[18] which actually enables one to determine E_x from its functional derivative, v_x:

$$E_x[n] = - \int n(\vec{r}) \vec{r} \cdot \vec{\nabla} v_x([n];\vec{r}) d^3r \qquad (48)$$

Levy and Perdew[18] derived Eq. (48) by utilizing the adiabatic connection technique[33-35] (here the electron-electron repulsion coupling constant was taken to the zero limit) together with the developments of Averill and Painter,[36] which in turn, followed the development of Slater for Xα theory.[37] A simpler proof is now given.
Following Ghosh and Parr,[38] for λ close to unity the

definition of v_x implies

$$E_x[n_\lambda] - E_x[n] = \int v_x([n];\vec{r})(n_\lambda - n)d^3r \tag{49}$$

to lowest order in $\delta n_\lambda = n_\lambda - n$. Next divide Eq. (49) by $\lambda - 1$ and let $\lambda \to 1$. This gives

$$(\partial E_x[n_\lambda]/\partial\lambda)_{\lambda=1} = \int v_x([n];\vec{r})(\partial n_\lambda/\partial\lambda)_{\lambda=1}d^3r \tag{50}$$

or

$$(\partial E_x[n_\lambda]/\partial\lambda)_{\lambda=1} = -\int n(\vec{r})\vec{r}\cdot\vec{\nabla}v_x([n];\vec{r})d^3r \tag{51}$$

which is the Ghosh-Parr relation.[38] Now, Levy and Perdew proved[18] Eq. (21), $E_x[n_\lambda] = \lambda E_x[n]$. Combination of Eqs. (21) and (51) proves Eq. (48). (It is important to note here that Eq. (48) is specific to the definition of exchange as given by Eq. (12) which employs Φ_n^{min} and not Φ_n^{HF}. In contrast, Eq. (48) is <u>not satisfied</u> by E_x^{HF} and v_x^{HF} because $E_x^{HF}[n_\lambda] \neq \lambda E_x^{HF}[n]$.)

8. ORBITAL-GENERATED KOHN-SHAM EXCHANGE-CORRELATION POTENTIAL.

The Kohn-Sham procedure[14] is the most practical one at present for quantitative calculations. Since orbitals are employed in this procedure anyway, and not just the density, there has been recent interest in the employment of orbital-generated exchange-correlation potentials where the Kohn-Sham equation takes the form

$$[-\tfrac{1}{2}\nabla^2 + v(\vec{r}) + u([n];\vec{r}) + v_{xc}(\{\phi_i\};\vec{r})]\phi_i = \epsilon_i\phi_i; \quad i = 1, 2, \ldots, N \tag{52}$$

where $\{\phi_i\} \to n$ and $u = \delta U/\delta n$. Upon self-consistency, $v_{xc}(\{\phi_i\};\vec{r})$ becomes $v_{xc}([n_o];\vec{r})$, where $v_{xc}([n];\vec{r})$ is the customary Kohn-Sham exchange-correlation potential and n_o is the ground-state density associated with v.

The next section shall discuss approximations to $v_x(\{\phi_i\};\vec{r})$, the exchange component of $v_{xc}(\{\phi_i\};\vec{r})$.

8.1. Theorem For Exact Exchange Component.

Ou-Yang and Levy[39] have recently proved the following theorem which should prove useful for helping to approximate $v_x(\{\phi_i\};\vec{r})$.

Consider a local exchange potential $\tilde{v}_x(\{\phi_i\};\vec{r})$. Then $\tilde{v}_x([n];\vec{r}) = \tilde{v}_x(\{\phi_i^{KS}\};\vec{r})$ is the exact Kohn-Sham exchange potential, $v_x([n];\vec{r}) = \delta E_x[n]/\delta n(\vec{r})$, if [1]: $\tilde{v}_x$ satisfies Eq. (48),

$$E_x[n] = -\int d^3r\, n(\vec{r})\vec{r}\cdot\vec{\nabla}v_x([n];\vec{r}) \tag{53}$$

where

$$E_x[n] = \langle\Phi_n^{SD}\{\phi_i^{KS}\}|V_{ee}|\Phi_n^{SD}\{\phi_i^{KS}\}\rangle - U[n]; \tag{54}$$

[2]: $\tilde{v}_x$ satisfies

$$\tilde{v}_x([n_\lambda];\vec{r}) = \lambda\tilde{v}_x([n];\lambda\vec{r}); \tag{55}$$

and [3]: $\bar{v}_x$ is a functional derivative.

Here ϕ_i^{KS} signifies a Kohn-Sham orbital and $\Phi_n^{SD}\{\phi_i^{KS}\}$ is the corresponding Kohn-Sham single determinant. (Note that $\Phi_n^{SD}\{\phi_i^{KS}\} = \Phi_n^{min}$ in those prevalent cases where Φ_n^{min} is a single determinant. This identity shall be assumed in the present discussion.)

The original Slater potential is $v_x^{Slater}(\{\phi_i\};\vec{r}) = \int n_x(\vec{r},\vec{r}')|\vec{r}-\vec{r}'|^{-1}d^3r'$, where $n_x(\vec{r},\vec{r}')$ is the exchange hole at $\vec{r}'$. The Slater potential satisfies condition [2] but definitely does not satisfy condition [1] and most likely does not satisfy condition [3]. The LDA, $cn^{1/3}$, satisfies conditions [2] and [3] but not [1].

Recently, as a possible correction to the Slater potential, Ou-Yang and Levy[39] have put forth the following potential, h_x^γ:

$$h_x^\gamma(\{\phi_i\};\vec{r}) - v_x^{Slater}(\{\phi_i\};\vec{r})$$

$$-(1-\gamma[n])\int_\infty^r dr' \int d^3r'' \frac{\partial}{\partial r'} \frac{n_x(\vec{s},\vec{r}'')}{|\vec{s}-\vec{r}''|} \qquad (56)$$

$$+\ \gamma[n] \left(\frac{n(\vec{r})}{3n(\vec{r})+\vec{r}\cdot\vec{\nabla}n(\vec{r})}\right)\int d^3r' \frac{\vec{r}\cdot\vec{\nabla}n_x(\vec{r},\vec{r}')}{|\vec{r}-\vec{r}'|}$$

where $\vec{s}=(r',\theta,\phi)$, and where $\gamma[n]$ is an arbitrary constant. Equation (56) was stimulated by the Harbola-Sahni potential and the above theorem.

The above potential satisfies conditions [1] and [2], decays correctly[40] as $-r^{-1}$, and is exact for one and two electrons. It is unlikely, however, that h_x^γ satisfies conditions [3].

With γ=o, h_x^γ becomes the Harbola-Sahni potential,[41] w_x,

for spherical densities. The work of Harbola and Sahni has been responsible, in large part, for rekindling interest in improving the original orbital-generated Slater potential.[32] It has been demonstrated that w_x yields highest-occupied orbital energies that are far superior[42] to those determined by the LDA exchange, and it has been shown by Wang, et al[43] that w_x is exact in the uniform limit, but that w_x differs from v_x in higher order terms.

The third condition is difficult to test. Hence, it is best to establish as many necessary conditions as possible. With this in mind, let's investigate in what ways h_x^γ acts like v_x because h_x^γ satisfies conditions [1] and [2] and in what ways h_x^γ might not behave like v_x even though h_x^γ satisfies conditions [1] and [2]. Well, from Eq. (21), condition [1], and Eqs. (48) - (51), it is clear that h_x^γ behaves like v_x with respect to a variation in n due to coordinate scaling. That is[44]

$$E_x[n_\lambda]-E_x[n]=\int h_x^\gamma([n];\vec{r})\delta n_\lambda d^3r \qquad (57)$$

to lowest order in $\delta n_\lambda = n_\lambda - n$. It can also be shown that the satisfaction of condition [2] guarantees[39] that h_x^γ is not the functional derivative of $G[n] = E_x[n] + B[n]$, where $B[n_\lambda] = B[n]$. (The functional derivative of G satisfies condition [1] but not condition [2].) On the other hand, the satisfaction of only conditions [1] and [2] obviously leaves open the possibility that h_x^γ does not behave like v_x with respect to variations which do not involve coordinate scaling of the density. So, it would be worthwhile to investigate non-scaling variations to derive additional necessary conditions.

Additional fundamental conditions arise from those variations which leave E_x unchanged.[44] For instance, consider a translation or a rotation of the density. For each of these $\delta E_x = 0$ by the homogeneity and isotropy of space. By employing a development which is somewhat analogous to the one leading to Eqs. (1) and (2) of Ref. (18) we have obtained[44]

$$\int d^3r v_x([n];\vec{r})\vec{\nabla}n(\vec{r})=0 \tag{58}$$

and

$$\int d^3r v_x([n];\vec{r})\vec{r}\times\vec{\nabla}n(\vec{r})=0. \tag{59}$$

The potential v_x^{Slater} has been tested[44] with respect to Eq. (58) and v_x^{Slater} has been found not to generally satisfy Eq. (58).

9. FUTURE CONSIDERATIONS: COORDINATE SCALING EQUALITY REQUIREMENTS FOR THE CORRELATION HOLE

I believe that future research will focus heavily upon the correlation energy functional for both inclusion in a full density-functional calculation or for inclusion as a tack-on to a Hartree-Fock calculation. It is my feeling that coordinate scaling requirements are stringent enough to dictate much about the form of E_c, especially for systems where the density is relatively inhomogeneous. Consequently, I conclude this chapter by asserting recently derived[24] scaling requirements for the correlation hole for use in the following well-known adiabatic connection formula[33-35] for E_c:

$$E_c[n] = \tfrac{1}{2}\int r_{12}^{-1}\, n(\vec{r}_1)\bar{\rho}_c([n];\vec{r}_1,\vec{r}_2)d^3r_1 d^3r_2 \tag{60}$$

where

$$\bar{\rho}_c([n];\vec{r}_1,\vec{r}_2) = \int_0^1 \rho_c([n,\lambda];\vec{r}_1,\vec{r}_2)d\lambda. \tag{61}$$

Here $\rho_c([n,\lambda];\vec{r}_1,\vec{r}_2)$ is the correlation hole of $\Psi_n^{\min,\lambda}$. (An attractive feature of Eq. (60) is that it implicitly includes the kinetic contributions to E_c). Eq. (60) is the particular adiabatic connection form of Langreth and Perdew[33] and Gunnarsson and Lundqvist.[34]

The correlation hole ρ_c is the unknown that has to be modeled. With this in mind, it is already known, of course, that it satisfies[45,46]

$$\int \rho_c([n,\lambda];\vec{r}_1,\vec{r}_2)d^3r_2 = 0 \tag{62}$$

for arbitrary n and arbitrary λ. I shall now reveal scaling equalities,[24] involving $\rho_c[n,\lambda]$, which are satisfied by the exact E_c and should help us obtain approximations to it. Perhaps the most important new equality is[24]

$$\rho_c([n,\lambda];\vec{r}_1,\vec{r}_2) = \lambda^3 \rho_c([n_\alpha,1];\lambda\vec{r}_1,\lambda\vec{r}_2) \tag{63}$$

With the latter expression, Eq. (60) becomes

$$E_c[n] = -\tfrac{1}{2}\int_0^1 \lambda \int r_{12}^{-1} n(\vec{r}_1)\rho_c([n_\alpha,1];\vec{r}_1\vec{r}_2)d^3r_1 d^3r_2 d\lambda \tag{64}$$

Equation (63) follows from [24] $\Psi_n^{min,\lambda}(\vec{r}_1,\ldots,\vec{r}_N) = \lambda^{3N/2}\Psi_{n_\alpha}^{min}(\lambda\vec{r}_1,\ldots,\lambda\vec{r}_N)$ which arises as a corollary to the "coordinate scaling theorem" of Levy and Perdew[18] (See Section 5). Further, from Ref. 18 and the definition of ρ_c, Eq. (63) implies[24]

$$\lim_{\lambda\to o} \lambda^3\rho_c([n_\alpha,1];\lambda\vec{r}_1,\lambda\vec{r}_2) = 0, \tag{65}$$

and $\lambda\int r_{12}^{-1}n(\vec{r}_1)\rho_c([n_\alpha,1];\vec{r}_1\vec{r}_2)d^3r_1d^3r_2$, which is never positive and decreases monotonically[24] with increasing λ, satisfies[24]

$$\lim_{\lambda\to\infty} \lambda\int r_{12}^{-1}n(\vec{r}_1)\rho_c([n_\alpha,1];r_1,r_2)d^3r_1,d^3r_2 \geq$$

$$-U[n]-E_x[n] \geq -U[n] \tag{66}$$

and thus decreases to a finite negative constant.

The 2-matrix of $\Psi_n^{min,\lambda}$, $\Gamma_{n,\lambda}(x_1'x_2'|x_1x_2)$, is conveniently partitioned as

$$\Gamma_{n,\lambda} = \Gamma_n^{KS} + \Gamma_{n,\lambda}^c \tag{67}$$

where Γ_n^{KS} is the 2-matrix of the Kohn Sham single determinant, Φ_n^{min}, and $\Gamma_{n,\lambda}^c$ is the correlation contribution to $\Gamma_{n,\lambda}$, which is related to $\rho_c[n,\lambda]$ by

$$\rho_c([n,\lambda];\vec{r}_1,\vec{r}_2) = 2\Gamma_{n,\lambda}^c(\vec{r}_1\vec{r}_2|\vec{r}_1\vec{r}_2)/n(\vec{r}_1). \tag{68}$$

Finally, from Eqs. (13), (64), and (68), we have

$$2(1\text{-}N)^{-1}\int\int\nabla^2\Gamma^c_{n,1}(x'x_2|xx_2)dx_2dx +$$

$$\int r_{12}^{-1}n(\vec{r}_1)\rho_c([n,1];\vec{r}_1,\vec{r}_2)d^3r_1d^3r_2 -$$

$$\int_o^1\lambda d\lambda\int r_{12}^{-1}\, n(\vec{r}_1)\rho_c([n_\alpha,1];\vec{r}_1,\vec{r}_2)d^3r_1d^3r_2 \qquad (69)$$

Equation (69) should prove a strong constraint if one wishes to model the Γ^c in terms of the density.

References

1. P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
2. L. H Thomas, Proc. Camb. Phil. Soc. 23, 542 (1927); E. Fermi, Rend. Acad. Naz. Lincei 6, 602 (1927).
3. M. Levy, Proc. Natl. Acad. Sci. (USA) 76, 6062 (1979).
4. M. Levy, Bull. Amer. Phys. Soc. 24, 626 (1979).
5. M. Levy, Phys. Rev. A 26, 1200 (1982).
6. P. W. Payne, J. Chem. Phys. 71, 490 (1979).
7. R. A. Harris and L. R. Pratt, J. Chem. Phys. 83, 4024 (1985).
8. M. Levy, pgs. 479-498 in Density Matrices and Density Functionals, R. Erdahl and V. H. Smith, Jr. (eds.), D. Reidel Publishing Company, Boston (1987); Proceedings of the A. John Coleman Symposium, Queen's University, Kingston, Canada, August, 1985.
9. J. E. Harriman, Phys. Rev. A 24, 680 (1981).
10. E. H. Lieb, Int. J. Quantum Chem. 24, 243 (1983).
11. S. M. Valone, J. Chem. Phys. 73, 4653 (1980).
12. J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Jr., Phys. Rev. Lett. 49, 1697 (1982).
13. J. K. Percus, Int. J. Quantum Chem. 13, 89 (1978).
14. W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
15. V. Sahni, J. Gruenbaum, and J. P. Perdew, Phys. Rev. B 26, 4371 (1982).
16. J. R. Sabin and S. B. Trickey, in Local Density Approximations in Quantum Chemistry and Solid State Physics, J. P. Dahl and J. Avery, (eds.), Plenum, New York (1984).
17. M. Levy, Phys. Rev., submitted for publication.
18. M. Levy and J. P. Perdew, Phys. Rev. A 32, 2010 (1985).
19. M. Levy, W. Yang, and R. G. Parr, J. Chem. Phys. 83, 2334 (1985).

21. M. Levy, J. P. Perdew, and V. Sahni, Phys. Rev. A. 30, 2745 (1984).
22. C. Herring, Phys. Rev. A 34, 2614 (1986).
23. J. P. Perdew and Y. Wang, Phys. Rev. B 33, 8800 (1986).
24. M. Levy, Phys. Rev., submitted for publication.
25. M. Levy, Int. J. Quantum Chem. S23, 000 (1989).
26. For recent reviews on X-ray diffraction accuracy see R. H. Blessing, Cryst. Rev. 1, 3 (1987) and P. Seiler, in Accurate Molecular Structures (eds. A. Domenicano, I. Hargittai, and P. Murray-Rust), Oxford University Press (1989).
27. J. M. Zuo, J. C. H. Spence, and M. O'Keeffe, Phys. Rev. Lett. 61, 353 (1988).
28. M. Levy and J. A. Goldstein, Phys. Rev. B 35, 7887 (1987).
29. J. P. Perdew, M. Levy, G. S. Painter, S. Wei, and J. B. Lagowski, Phys. Rev. B 37, 838 (1988).
30. F. Aryasetiawan and M. J. Stott, Phys. Rev. B 38, 2974 (1988).
31. J. Cioslowski, Phys. Rev. Lett. 60, 2141 (1988).
32. J. C. Slater, Phys. Rev. 81, 385 (1951).
33. D. C. Langreth and J. P. Perdew, Solid State Commun., 17, 1425 (1975) Phys. Rev. B 15, 2884 (1977).
34. O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B 13, 4274 (1976).
35. J. Harris and R. O. Jones, J. Phys. F 4, 1170 (1974).
36. F. W. Averill and G. S. Painter, Phys. Rev. B 24, 6795 (1981).
37. J. C. Slater, J. Chem. Phys. 57, 2389 (1972).
38. S. K. Ghosh and R. G. Parr, J. Chem. Phys. 82, 3307 (1985).
39. H. Ou-Yang and M. Levy, submitted to Phys. Rev.
40. J. D. Talman and W. F. Shadwick, Phys. Rev. A 14, 36 (1976).
41. M. K. Harbola and V. Sahni, Phys. Rev. Lett. 62, 489 (1989).
42. Y. Li, M. K. Harbola, J. B. Krieger, and V. Sahni, Phys. Rev. A, in press.
43. Y. Wang, J. P. Perdew, J. A. Chevary, L. P. Macdonald, and S. H. Vosko, Phys. Rev. A, in press.
44. M. Levy and H. Ou-Yang, unpublished.
45. R. G. Parr and W. Yang, Density-Functional Theory of Atoms and Molecules, Oxford University Press, New York (1989).
46. R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. 61, 689 (1989).

DENSITY-FUNCTIONAL THEORY OF THE BAND GAP

M. Schlüter

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

L. J. Sham

Department of Physics
University of California–San Diego
La Jolla, California 92093

1. INTRODUCTION

The density-functional theory /1,2/ (DFT) gives a framework for the interaction effects in an inhomogeneous system which, in the simple local-density approximation (LDA), has proven to be quantitatively rather good for many ground-state properties /3/. The one-electron Schrödinger equation with the effective LDA potential derived from the DFT yields wave functions which form a quite accurate ground-state density. This is the basis of the good approximation for the ground-state properties. The eigenvalues from the same equation do not, however, yield accurate band gaps for semiconductors and insulators. Errors of the order or larger than 50% are not uncommon.

Sham and Kohn /4/ pointed out that the eigenvalues of the DFT one-particle Schrödinger equation (even with the exact density-functional potential in place of the LDA) do not correspond to the quasi-particle energies except at the Fermi level. They gave an LDA for the self-energy as a functional of the density. From the self-energy, the poles of the Green's function can be determined to yield the quasi-particle energies. Pickett and Wang /5/ have recently applied this method to the evaluation of the band gaps. Since then, much progress has been made to construct the self-energy with the LDA orbitals but without the LDA on the self-energy itself /6,7/.

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

By construction, the highest eigenvalue of the DFT equation corresponds to the energy of the highest occupied quasi-particle state. This fact can be used to provide a firm theoretical basis for the problem of the fundamental band gap, even though the approach does not apply rigorously to the general band structure. In a semiconductor or insulator, the band gap is rigorously defined as the difference between the lowest conduction-band energy and the highest valence-band energy, which, being the smallest energies to add and remove, respectively, an electron from the N-electron system, can then be shown to correspond, respectively, to the highest occupied DFT eigenvalues of the $N + 1$ and N particle states.

Emphasizing the dependence of the density functionals on the number of particles brings out the difference between the highest occupied energy of the $(N+1)$-particle system and the $(N+1)$th level of the N-particle system. The difference is easily shown to be the discontinuity Δ in the DFT potential when the number of particles is increased by one /8,9/. The gap is then composed of two terms, the DFT gap, defined from the eigenvalues of the N-particle DFT equation, and the discontinuity Δ. The discrepancy between the LDA gap and the measured gap has two components: the improvement of the N-particle DFT energy values over LDA and the discontinuity. The relative importance of these two contributions is of central interest here.

We have initially studied these questions by finding field-theoretic expressions for the exact exchange-correlation potential /10/ and for the discontinuity /11/, and by evaluating these expressions for a two-plane wave model for an insulator both in the exchange-only approximation /11/ and including correlation /12/. In subsequent collaboration with Godby /7/, we calculated the exchange-correlation potential $V_{xc}(r)$ and the discontinuity Δ for semiconductors and insulators from the self-energy operator $\Sigma(r, r', \omega)$ within the random phase approximation. In this chapter, we review some of the results.

2. THE CONNECTIONS BETWEEN QUASIPARTICLE ENERGIES AND DF THEORY

There is a formal resemblance of Dyson's equation, which gives the quasiparticle energies, to the Schrödinger equation for the effective-one-electron eigenvalues of DFT: each is a Schrödinger equation for fictitious, noninteracting electrons moving in a effective potential (although in one case this "potential" – the self energy – is nonlocal and energy dependent). This does not mean in general that there is a relation between the quasiparticle energies E_i and the DFT eigenvalues $E_{i,\mathrm{DFT}}$.

However, the highest occupied density functional eigenenergy does represent the chemical potential in the conductor case and the valence band edge in the insulator or semiconductor case. The band gap of an insulator or a semiconductor can be defined precisely in terms of the ground state energy as a function $E_{tot}^{(M)}$ of the number of particles M. If the insulating ground state has N particles, the conduction band edge is the change of the total ground state energy when an electron is added and the valence band edge is given by the change when an electron is removed:

$$E_c \equiv E_{tot}^{(N+1)} - E_{tot}^{(N)} \tag{1}$$

$$E_v \equiv E_{tot}^{(N)} - E_{tot}^{(N-1)} .$$

The band gap is naturally the difference:

$$E_g = E_c - E_v . \tag{2}$$

It is straightforward to show /11/ with the help of the variational theorem that the highest DFT eigenvalue is given by the energy to remove an electron:

$$E_{M,\mathrm{DFT}}^{(M)} = E_{tot}^{(M)} - E_{tot}^{(M-1)} . \tag{3}$$

Now, the Hohenberg-Kohn theorem is implicitly valid for a fixed number of electrons, M. The density function equation and the exchange-correlation potential in it, V_{xc}, are implicitly defined as functions of M. Since $M=N$ gives the insulating ground state, when an electron is removed, the highest occupied state is not changed much (only to $O(1/N)$) and $V_{\mathrm{xc}}^{(N-1)}$ is the same as $V_{\mathrm{xc}}^{(N)}$ but that when an electron is added across the gap, the $(N+1)$th state is very different and there is a discontinuity in $V_{\mathrm{xc}}^{(N)}$ as N is changed to $N+1$:

$$V_{xc}^{(N+1)} = V_{xc}^{(N)} + \Delta, \tag{4}$$

where Δ is independent of position since the electron density is changed only to $O(1/N)$. It follows that the difference between the true gap, Eq. (2) and the density functional gap given by

$$E_{g,DFT} = E_{N+1,DFT}^{(N)} - E_{N,DFT}^{(N)} \tag{5}$$

is just the potential discontinuity, Δ:

$$E_g = E_{g,\mathrm{DFT}} + \Delta. \tag{6}$$

The situation is depicted diagrammatically in Fig. 1.

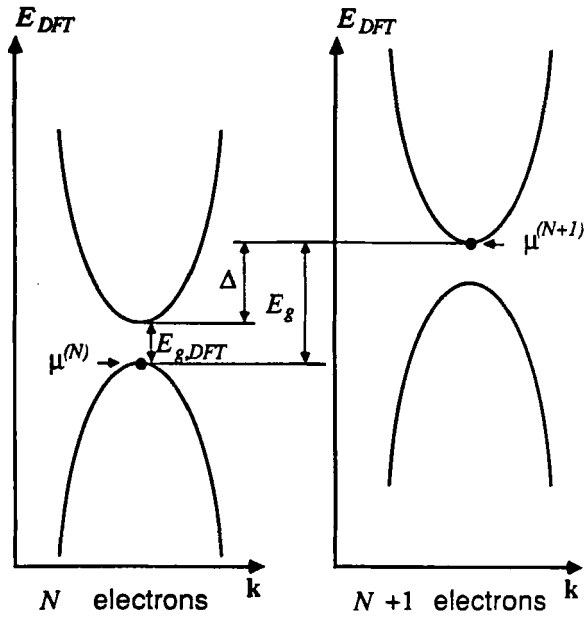


Fig. 1. Illustration of the significance of Δ, the discontinuity in V_{xc}. The exact DFT Kohn-Sham one-electron energies are shown in the form of a band structure for the N- and $(N+1)$- particle system. The two differ in a uniform increase of the eigenvalues by Δ, as explained in the text. The quasiparticle gap E_g is the difference between the two eigenvalues indicated: $E_g = E_{N+1,\mathrm{DFT}}^{(N+1)} - E_{N,\mathrm{DFT}}^{(N)}$. It is evident that $E_g = E_{g,\mathrm{DFT}} + \Delta$.

Δ has been shown to be a significant fraction of the energy gap for a one-dimensional model semiconductor within a two-plane wave basis set /11,12/, and more recently also for real semiconductors /7/. This is important. If Δ is small, exact DFT will give accurate band structures, and attention should be paid to going beyond the LDA in density-functional theory; if Δ is large, no attempt to add nonlocal-density corrections to the LDA will improve the calculated DFT band structures, and one must either go outside DFT, or take the discontinuity in V_{xc} into account explicitly as a correction to the DFT band structure when quasiparticle energies are needed. Thus Δ lies at the heart of the relationship between the quasiparticles and DFT for semiconductors and insulators. In a periodic solid, where the momentum $\mathbf{k}$ is a good quantum number for both the DFT and quasiparticle wave functions, the rigorous extension $\Delta(\mathbf{k})$ is possible to all the minimum direct band gaps throughout the Brillouin zone with energies below the Auger threshold.

A practical method /9,10,11/ to calculate an (approximate, but non-LDA) exchange correlation potential $V_{xc}(\mathbf{r})$ and therefore Δ is by way of solving Dyson's equation first for a given (approximate) self-energy $\Sigma(r, r', \omega)$ and then by requesting that both the DFT potential V_{xc} and the self-energy Σ produce the same total electron density. This leads to the integral equation, implicitly defining V_{xc} for a known Σ:

$$\mathrm{Im} \int_{-\infty}^{\mu} [G_{\mathrm{DFT}}(\Sigma - V_{xc})G]_{r,r} d\omega = 0 , \tag{7}$$

where the one particle Green's function G_{DFT} contains V_{xc} and G contains Σ.

3. CALCULATIONAL DETAILS

We begin the calculation of Δ with a well-converged LDA calculation for the material at its experimental lattice constant, using a plane-wave basis set and nonlocal pseudopotentials /7/. The whole calculation is performed within this plane-wave basis. (The different stages in the calculations, starting from the LDA band structure, are summarized as a flow diagram in Fig. 2.) Next, the LDA wave functions and energies are used to form the random-phase

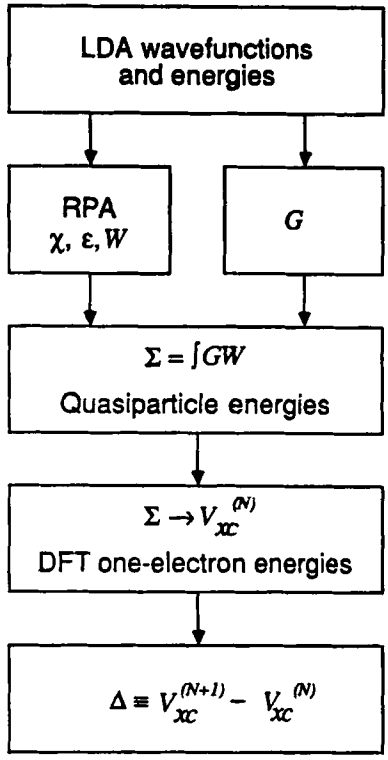


Fig. 2. Flow diagram of the calculation, starting from the self-consistent LDA pseudopotential calculation, forming the screened Coulomb interaction W and the Green's function G, combining them to form Σ and then using Σ to obtain the quasiparticle energies E_i and also the true exchange-correlation potential V_{xc}.

approximation (RPA) dielectric function, $\varepsilon(\mathbf{r},\mathbf{r}',\omega)$. Because of the periodicity of the crystal, ε (and all nonlocal quantities considered here) can be Fourier transformed, yielding $\varepsilon_{GG'}(\mathbf{k},\omega)$. The inclusion of the off-diagonal terms in the dielectric matrix $\varepsilon_{GG'}$, corresponds to taking account of the so-called *local-field effects*, since a purely diagonal dielectric matrix would imply that ε depended only on $|\mathbf{r}-\mathbf{r}'|$, thereby ignoring the fact that the unit cell is not homogeneous but can respond to a perturbation by setting up locally varying fields. The local-field terms are crucial (though mainly in their effect on the

strength of the screening hole rather than its shape) and contribute directly to the differing strengths of the self-energy operator at different points in the unit cell and therefore to the band-gap correction.

Despite the simplifications of periodicity, the evaluation of ε is still very demanding computationally, and only in the last few years have calculations of $\varepsilon_{GG'}(\mathbf{k},\omega)$ been published /13/, and then only for a very limited range of $\mathbf{G},\mathbf{G}'$, $\mathbf{k}$, or ω. We achieved /7/ a considerable reduction in computation while maintaining full accuracy by calculating ε for imaginary, rather than real, ω, using the analytic continuation of the RPA expression to imaginary frequencies, so that there is no need to take elaborate precaution for handling the poles of ε that lie just off the real ω axis. Nevertheless, the calculation of the dielectric function represents a major computational effort in the whole work and simplifications without sacrificing accuracy should prove useful for future work /14,15/.

We then form the *inverse dielectric function* ε^{-1} and the *screened Coulomb interaction* $W(\mathbf{r},\mathbf{r}',\omega) = \varepsilon^{-1}V$, again in reciprocal space, by straightforward matrix inversion and multiplication for each $\mathbf{k}$. W is then used in the *GW* approximation /16/ for the self-energy:

$$\Sigma(\mathbf{r},\mathbf{r}',\omega) = \frac{i}{2\pi}\int_{-\infty}^{\infty} e^{i\omega\delta}W(\mathbf{r},\mathbf{r}')G(\mathbf{r},\mathbf{r}',\omega+\omega')d\omega' , \tag{8}$$

which is the leading term in an expansion of Σ in powers of the relatively weak screened Coulomb interaction (as opposed to the strong bare Coulomb interaction) and corresponds to setting the so-called vertex function Γ to a δ function. The remaining terms in the series, the *vertex corrections*, correspond to successive corrections to the vertex function. Calculations of Γ have so far been restricted to the electron gas case.

The calculated self-energy is then used in two ways. First, the quasiparticle energies E are found, using e.g. second-order perturbation theory in $\Sigma - V_{xc}^{LDA}$. Second, we use Σ to calculate a DFT exchange-correlation potential, V_{xc}, for the semiconductor, using the relationship between Σ and the DFT V_{xc} given in eqn. (7). In order to make this tractable, we replace both G and G_{DFT} by G_{DFT}^{LDA}. Tests indicate that this is of little consequence in the final V_{xc}.

4. RESULTS

Table I gives the calculated quasiparticle energies for silicon and, for comparison, the LDA eigenvalues and the experimental values. The DFT eigenvalues with the calculated V_{xc} are also given, and will be discussed below. Table II, III, and IV are the corresponding results for gallium arsenide, aluminum arsenide, and diamond, respectively /7/.

The general trends are immediately obvious, and are the same for each of the four materials: The quasiparticle energies calculated using the *GW* self-energies are in excellent agreement with experiment (generally within 0.1 eV), and differ from the LDA eigenvalues mainly in a very approximately rigid vertical displacement of the conduction bands relative to the valence bands. The neglect of vertex corrections does not seem to be crucial. A detailed discussion of each material is given in ref. 7.

TABLE I. Quasiparticle energies (*GW*) at Γ and L, and the minimum gap in silicon, in eV. The experimental (Expt.) energies (at $T = 0$ where known), and the DFT eigenvalues using (i) V_{xc}^{LDA} ("LDA") and (ii) the V_{xc} calculated from Σ ("DFT") are also shown. For details see ref. 7.

	LDA	DFT	*GW*	Expt.	*GW* − DFT
Γ'_{25v}	−0.07	0.00	0.00	0.00	0.00
L_{2v}	−1.29	−1.21	−1.19	−1.2±0.2, −1.5	0.02
c-band minimum	0.45	0.66	1.24	1.17	0.58
Γ_{15c}	2.50	2.68	3.30	3.40	0.62
Γ_{2c}	3.49	3.66	4.27	4.19	0.61
L_{1c}	1.46	1.62	2.30	2.1, 2.4	0.68
L_{3c}	3.30	3.49	4.11	4.3±0.2, 4.0	0.62
Minimum gap	0.52	0.66	1.24	1.17	0.58

TABLE II. Quasiparticle and other energies in GaAs. The spin-order splitting shown in parentheses is removed. (See Table I for details.)

	LDA	DFT	*GW*	Expt.	*GW* − DFT
Γ_{15v}	−0.13	0.00	0.00	0.00 (+0.11, −0.22)	0.00
X_{5v}	−2.70	−2.55	−2.64	−2.80 (+0.05, −0.02)	−0.09
L_{3v}	−1.20	−1.06	−1.11	−1.30 (±0.05)	−0.05
Γ_{1c}	0.54	0.91	1.58	1.63	0.67
X_{1c}	1.36	1.59	2.19	2.09	0.60
X_{3c}	1.53	1.74	2.41	2.49	0.67
L_{1c}	1.02	1.30	1.93	1.93	0.63
Minimum gap	0.67	0.91	1.58	1.63	0.67

TABLE III. Quasiparticle and other energies in AlAs. (See Tables I, II for details)

	LDA	DFT	GW	Expt.	GW − DFT
Γ_{15v}	−0.04	0.00	0.00	0.00 (+0.09, −0.18)	0.00
X_{5v}	−2.21	−2.17	.30	−2.32 (+0.12, −0.06)	−0.13
L_{3v}	−0.85	−0.82	−0.89(=L_{1c}−4.01)	L_{1c} − 4.00	−0.07
Γ_{1c}	2.38	2.56	3.35	3.22	0.79
X_{1c}	1.33	1.55	2.18	2.32	0.63
L_{1c}	2.18	2.38	3.12	2.58	0.74
Minimum gap	1.37	1.55	2.18	2.32(+0.18, −0.09)	0.63

TABLE IV. Quasiparticle and other energies for diamond. (See Table I for details.)

	LDA	DFT	GW	Expt.	GW − DFT
Γ'_{25v}	−0.05	0.00	0.00	0.00	0.00
X_{4v}	−6.31	−6.26	−6.71		−0.45
L'_{3v}	−2.79	−2.79	−2.98		−0.19
c-band minimum	3.85	4.21	5.33	5.48	1.12
Γ_{15c}	5.57	5.72	7.26	7.12	1.54
Γ_{2c}	13.11	13.24	14.83	15.3±0.5	1.59
X_{1c}	4.53	4.81	5.84(=X_{4v}+12.55)	X_{4v}+12.55	1.03
L_{3c}	8.38	8.48	9.63(=L'_{3v}+12.6.1)	L'_{3v} + 12.0	1.15
L_{1c}	8.68	8.86	10.55		1.69
Minimum gap	3.90	4.21	5.33	5.48	1.12

It has been known for some time that the main discrepancies between the LDA energies in silicon and the experimentally measured quasiparticle energies can be corrected by simply adding a constant (about 0.7 eV in Si) to the LDA energies above the band gap. This procedure is often referred to as a *scissors operator* /17/, because it can be thought of as cutting the band structure along the band gap and moving the conduction bands rigidly upwards. To study the range of applicability of the scissors operator we have plotted in Fig. 3 the difference between the quasiparticle and LDA energies as a function of **k** for the top valence band and bottom conduction band for each of the four

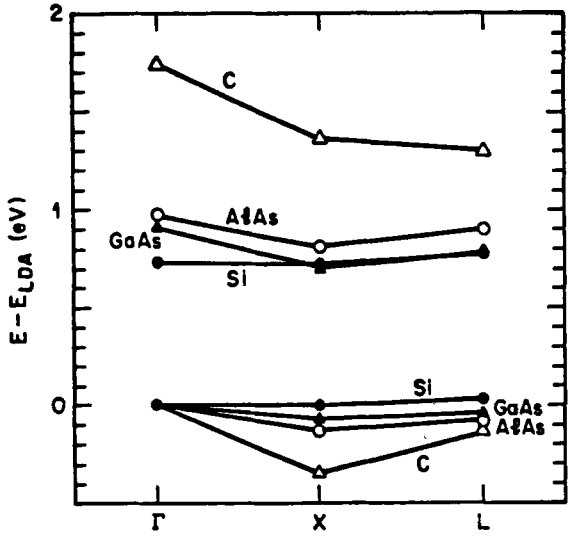


Fig. 3. The varying validity of the "scissors operator." The difference between the calculated quasiparticle energies and the LDA energies (relative to the valence-band maximum) in the lowest conduction band and highest valence band is plotted against **k** for each materials. If the scissors operator is valid, this quantity should be independent of **k**. It is evident that the scissors operator is accurate to about 0.1 eV in silicon, but only to 0.2, 0.2, and 0.4 eV, respectively, in GaAs, AlAs, and diamond.

materials. Clearly, if the scissors operator is to be valid the curves should be a **k**-independent constant for the conduction band, and zero in the valence band. It is evident that the scissors operator is indeed accurate to about 0.1 eV in silicon for the bands shown, but that its accuracy is substantially diminished for the other materials, particularly diamond.

In the process of evaluating the full *GW* operator, quasiparticle energies corresponding to other approximate self-energy operators can also easily be calculated: e.g. the bare exchange (Hartree-Fock) operator and the statically screened exchange operator each of which is energy independent. In each of the four materials the bare-exchange gaps and widths of the individual bands are much too large: the minimum gaps are approximately 5, 6, 7, and 12 eV in Si, GaAs, AlAs, and diamond. Screening with the static dielectric matrix reduces the gaps dramatically (to 0.4, 1.2, 1.2, and 4.7 eV), but does not fully correct for the lack of energy dependence. Dynamical screening, as included in the *GW* calculations, is thus necessary in calculating accurate quasiparticle energies.

The important way in which Σ is essentially different from V_{xc} is that Σ is a *nonlocal* operator, while V_{xc} is a *local* potential. This nonlocality is essential in determining the correct quasiparticle energies (and, in particular, the correct band gap). Despite the complexity of the calculation of Σ, the operator has a relatively simple structure. The operator is dominated by a single "hole," centered about $\mathbf{r} = \mathbf{r}'$ and approximately spherical. Moreover, the shape of the hole (though not its amplitude) is approximately independent of the location of $\mathbf{r}$. The range of this hole is seen to be approximately 2.0 r_s for the jellium of the corresponding average density; for example, 2.1 Å in silicon. This length is comparable with the wavelength of the wave functions near the band gap, thus providing further proof that the nonlocality of Σ is essential. To examine this point in more detail, we consider a tight-binding approximation to the band structure of silicon. The states in the highest valence band and lowest conduction band are of quite distinct character in this picture: At Γ, for example, the valence-band maximum is *bonding p*, while the lowest conduction band is *antibonding p*. Thus the conduction-band wave function have an extra node between the atoms and the effective wavelength in this important part of

the unit cell (where the self-energy operator has greatest strength) is smaller. The matrix elements $(\psi \mid \Sigma \mid \psi)$ are therefore less negative in the conduction band than they would be if Σ were of the same strength but local, and so the quasiparticle energies in the conduction bands are raised and the band gap increases. An increased band gap is therefore an immediate consequence of the essential nonlocality of Σ. This nonlocality extends over a characteristic length of order interatomic distance.

The DFT eigenvalues using the calculated (non-LDA) V_{xc} are also shown in Tables I-IV for each material, and in addition silicon's $V_{xc}(\mathbf{r})$ is compared with $V_{xc}^{LDA}(\mathbf{r})$ in Fig. 4. In each case the V_{xc} shown is that for the N-particle system – the semiconductor with its valence bands exactly filled. The corresponding potential and eigenvalues for the $(N+1)$ particle system would differ from those given by an $\mathbf{r}$-independent constant Δ; that is, by $E_g - E_{g,\mathrm{DFT}}$, where E_g is the calculated quasiparticle minimum gap and $E_{g,\mathrm{DFT}}$ is the gap when V_{xc} is used. The similarity of the two potentials in Fig. 4 is striking, especially when one considers their quite different origins, and the fact that they have not been aligned in energy.

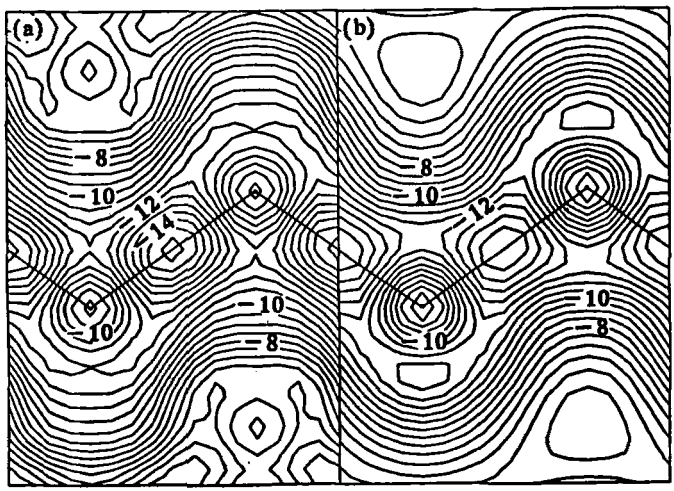


Fig. 4. The LDA and true exchange-correlation potentials, (a) V_{xc} and (b) V_{xc}^{LDA}, for silicon (the N-particle system), shown in eV in the (110) plane containing the bond chains as indicated.

The discontinuity Δ in V_{xc} upon addition of an electron to the N-electron system may be obtained by subtracting the calculated DFT minimum gap from the calculated *GW* minimum quasiparticle gap. Δ is shown in Table V and graphically in Fig. 5. Although Δ increases with the band gap, it is a remarkably constant proportion of the error in the LDA band gap, showing that even diamond is sufficiently locally free-electron-like for the LDA to be a good approximation to the true DFT potential. The fact that Δ is in each case a substantial proportion of the band gap shows that attempts to calculate improved quasiparticle energies within DFT by going beyond the LDA would bring no significant improvement.

It is interesting to compare this V_{xc} with previous attempts to go beyond the LDA. The most successful such potentials that have stayed within the general framework of DFT have been the weighted-density approximation /18/ (WDA), the average-density approximation /18/ (ADA), and gradient-expansion techniques /19/ (but only when explicitly designed to obey the important sum rule on the exchange-correlation hole). In each case the Kohn-Sham eigenvalues changed: the band gaps increased somewhat. In Table VI we compare some sample band gaps /20,21/ in the LDA, WDA, and with our V_{xc}. The general trend is the same with each of the three non-LDA V_{xc}'s: the band

TABLE V. Δ, the discontinuity in the exchange-correlation potential; E_g, the calculated minimum quasiparticle band gap; and the LDA gap error $E_g - E_{g,LDA}$ for the four materials.

Material	Δ (eV)	E_g (eV)	LDA gap error (eV)	Δ/(LDA gap error) (%)
Si	0.58	1.24	0.72	81
GaAs	0.67	1.58	0.91	74
AlAs	0.65	2.18	0.81	80
Diamond	1.12	5.33	1.43	78

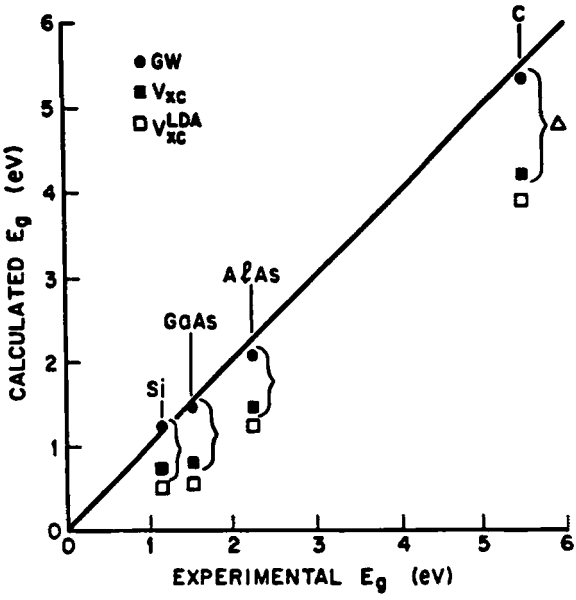


Fig. 5. The calculated minimum band gap (i) the *GW* approximation, (ii) DFT, and (iii) the LDA plotted against the experimental band gap. The 45° line is a guide to the eye. Δ, the discontinuity in the exchange-correlation potential, is indicated.

gaps are increased by between 0.05 and 0.34 eV, so that the 55% error in the LDA minimum gap in silicon, for example, is reduced to 26-43%. A detailed examination shows a surprising agreement between our V_{xc} and the WDA in silicon: they are within about 0.02 eV of each other. However, this excellent agreement is not obtained in GaAs. The reason for the agreement in silicon is unclear.

5. CONCLUSIONS

We have shown how the self-energy of a semiconductor may be calculated within the *GW* approximation and then used to calculate both the quasiparticle energies and a density-functional-theory exchange-correlation potential. In the four materials studied (Si, GaAs, AlAs, and diamond) the quasiparticle energies

TABLE VI. The direct band gaps and the minimum gap of Si and GaAs, in eV, in (i) experiment (Expt.) and (ii) the LDA, and the corrections to the LDA values in (iii) the weighted-density approximation (using a jellium exchange-correlation hole) (WDA), (ref. 20,21) (iv) the WDA using a Levine-Louie-model exchange-correlation hole (ref. 21) [WDA(LL)], and (v) the V_{xc} calculated here from the self-energy (DFT). See ref. 7 for details.

Gap	Expt.	LDA	Corrections to LDA WDA	WDA(LL)	DFT
Si					
Γ	3.40	2.57	+0.12	+0.24	+0.11
X	4.25	3.53	+0.10	+0.23	+0.11
L	3.3	2.75	+0.04	+0.13	+0.08
Minimum gap	1.17	0.52	+0.15	+0.34	+0.14
GaAs					
Γ (= minimum gap)	1.63	0.67	+0.06		+0.24
X	4.89	4.06	+0.13		+0.08
L	3.23	2.22	+0.08		+0.14

are in good agreement with reliable experimental data. However, the one-electron density-functional-theory eigenvalues corresponding to the calculated exchange-correlation potential are significantly different and are instead close to the eigenvalues obtained using the conventional local-density approximation for exchange and correlation. In particular, the DFT minimum band gap is substantially smaller than the quasiparticle gap, with the difference (equal to Δ, the discontinuity in the exchange-correlation potential upon addition of an extra electron) accounting for about 80% of the severe and well-known error in the LDA minimum gap. There are also considerable discrepancies in other parts of the energy spectrum. Thus, improving the density-functional-theory exchange-correlation potential beyond the local-density approximation will not lead to the correct quasiparticle energies. The *GW* method, which goes outside density-functional theory, is capable of giving good quasiparticle properties. On the other hand, the LDA is seen to be an even better approximation to the true DFT exchange-correlation potential than it was previously known to be.

REFERENCES

1. P. Hohenberg and W. Kohn (1964). Phys. Rev. **136**, B864.
2. W. Kohn and L. J. Sham (1965). Phys. Rev. **140**, A1133.
3. For reviews, see, for example, M. Schlüter and L. J. Sham (1982). Physics Today **35**(2), 30, or the book *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundquist and N. H. March (Plenum, New York, 1983).
4. L. J. Sham and W. Kohn (1966). Phys. Rev. *145*, 561.
5. W. E. Pickett and C. S. Wang (1984). Phys. Rev. B**30**, 4719.
6. M. S. Hybertsen and S. G. Louie (1985). Phys. Rev. Lett. **55**, 1418.
7. R. W. Godby, M. Schlüter and L. J. Sham (1988). Phys. Rev. **B37**, 10159 and references therein.
8. J. P. Perdew and M. Levy (1983). Phys. Rev. Lett. **51**, 1884.
9. L. J. Sham and M. Schlüter (1983). Phys. Rev. Lett. **51**, 1888.
10. L. J. Sham (1985). Phys. Rev. **B32**, 3876.
11. L. J. Sham and M. Schlüter (1985). Phys. Rev. **B32**, 3883.
12. M. Lannoo, M. Schlüter and L. J. Sham (1985). Phys. Rev. **B32**, 3890.
13. M. S. Hybertsen and S. G. Louie (1987). Phys. Rev. **B35**, 5585, **B35**, 5602, and references therein.
14. M. S. Hybertsen and S. G. Louie (1988). Phys. Rev. **B37**, 2733.
15. F. Gygi and A. Baldereschi (1989). Phys. Rev. Lett. **62**, 2160.
16. L. Hedin (1965). Phys. Rev. **139**, A796.
17. G. A. Baraff and M. Schlüter (1984). Phys. Rev. **B30**, 3460.
18. O. Gunnarsson, M. Jonson and B. I. Lundqvist (1979). Phys. Rev. **B20**, 3136.
19. D. C. Langreth and M. J. Mehl (1981). Phys. Rev. Lett. **47**, 446; Phys. Rev. **B 28**, 1809 (1983).
20. F. Manghi, G. Riegler, C. M. Bertoni, C. Calandra and G. B. Bachelet (1983). Phys. Rev. **B28**, 6157.
21. M. S. Hybertsen and S. G. Louie (1984). Phys. Rev. **B30**, 5777.

SIZE-CONSISTENCY, SELF-INTERACTION CORRECTION, AND DERIVATIVE DISCONTINUITY IN DENSITY FUNCTIONAL THEORY

John P. Perdew

Department of Physics and Quantum Theory Group
Tulane University
New Orleans, Louisiana
U.S.A. 70118

ABSTRACT

Three size-consistency principles of electronic structure theory are set forth. "Separability" and "extensivity" are respected even by local or semilocal density functional approximations, which succeed for that and other reasons. The strong tendency of a separated subsystem to reject fractional electron number is a consequence of the derivative discontinuity of the energy; this "integer preference" is probably not respected by any practical density functional approximation. However, all three size-consistency principles are obeyed by theories which construct the energy from localized orbitals. The self-interaction correction (SIC) to the local spin density approximation favors localized orbitals for most systems. If this fails for metals, then SIC gives rise to a "false surface energy". The self-interaction correction is reviewed, along with its exact-theory "doppelganger", the derivative discontinuity. The latter reveals the physical content of the exact Kohn-Sham orbital energies and resolves the "band-gap problem".

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

1. INTRODUCTION

The remarkable Hohenberg-Kohn theorem[1-4] is a mainstay of the physicist's or chemist's unified understanding of atoms, molecules and solids. Consider a system of electrons bound by an external potential $v(\vec{r})$, which typically arises from a density $n_+(\vec{r})$ of positive charge:

$$v(\vec{r}) = -\int d^3r' \frac{n_+(\vec{r}')}{|\vec{r}-\vec{r}'|} \quad . \tag{1}$$

The theorem asserts the existence of a functional $E_v[n]$ whose minimum yields the ground-state energy E and electron density $n(\vec{r})$. Useful approximations to this functional are often suggested by or measured against general principles such as "size-consistency".

Three of these principles are stated below. Only the first and half of the second (separability and volume extensivity) are conventional size-consistency principles[5,6] of electronic structure theory.

(1) Separability. For a system composed of two or more objects separated by great distances of empty space, the total energy is the sum of the energies of the separate objects. Each object might be an atom, for which the positive charge distribution $n_+(\vec{r})$ is that of the nucleus. Separable approximations are desirable whenever the energy of a molecule or solid must be calculated as a function of the nuclear positions, to ensure that the binding energy curves have proper limiting behavior. In a familiar example from quantum chemistry, the spin-restricted Hartree-Fock approximation is not separable for open-shell objects (e.g., two hydrogen atoms), and truncated configuration interaction is not separable for any choice of objects.[5,6]

(2) Extensivity. A solid is a large periodic array of positive charge $n_+(\vec{r})$ confined within a compact volume, and neutralized by the electron density $n(\vec{r})$. If the bulk unit cell is free of electric monopole and dipole moments, the total energy is

$$E = f\,V + \sigma\,A, \qquad (2)$$

where V and A are the volume and surface area of the solid. The bulk energy f depends upon the nature of the unit cell, while the surface energy[7] also depends upon which crystal planes are exposed at the surface (but not otherwise upon the shape of the solid). Atoms will cluster together if there is a unit cell which minimizes the bulk energy f; the clusters will tend to have compact, surface-minimizing shapes if the surface energy σ is positive.

(3) Integer preference. In a collection of separated objects, nature prefers to locate an integer number of electrons on each object, and will do so except in the presence of unusual degeneracies between two or more integer configurations. Approximations should reflect this preference, for the same reason that they should be separable. For example, consider a sodium atom separated by distance R from a chlorine atom.[5,8] If $R < 18$ bohr, the ground state is $Na^{+1}...Cl^{-1}$, in which sodium and chlorine assume closed-shell configurations with 10 and 18 electrons, respectively; these ionic configurations are stabilized by the energy of mutual Coulomb attraction. If $R > 18$ bohr, the ground state is Na...Cl, in which sodium and chlorine assume neutral configurations with 11 and 17 electrons, respectively. In fact, there is never any fractional electron number on an atom in an overall-neutral collection of well-separated atoms.

Two known limits of the density functional have proven to be very powerful:

(a) Slowly-varying limit. When the density varies slowly over space, the non-electrostatic pieces of the energy functional may be represented to asymptotic accuracy by a gradient expansion.[1] For example, the exchange-correlation energy is

$$E_{xc}[n] = \int d^3r\; n\varepsilon_{xc}(n) + \int d^3r\; C_{xc}(n)\;|\nabla n|^2/n^{4/3} + \ldots \qquad (3)$$

Often the local density approximation (first term on the right in Eq. (3)) suffices even for realistic density variations.

(b) One-electron limit. For a single electron, separated

from all others, the total energy is just the eigenvalue of the one-electron Schödinger equation with potential $v(\vec{r})$.

Local density approximations, gradient expansions, and generalized gradient approximations[9,10] are justly popular, not only because they are simple but also because they are manifestly separable and extensive. Hence they can be used with some confidence to calculate small energy changes of chemical and physical interest, even though such changes may be dwarfed by the errors these local or semilocal approximations make in the total energy. However, they are not exact in the one-electron limit, and fail to display integer preference (as described in section 2).

A "self-interaction correction" is a procedure to correct a density functional which is not exact in the one-electron limit. After a reprise of Hohenberg-Kohn-Sham theory[1,11] in section 2, the Perdew-Zunger (1981)[12] recipe for self-interaction correction (SIC) will be reviewed in section 3. When applied to the local spin density (LSD) approximation for the exchange-correlation energy, SIC is separable and displays integer preference. It is at least partly extensive, reproducing the volume term in Eq. (2). Is the surface term also reproduced for a metal? Is SIC exact in the uniform-density limit? These questions will be discussed in section 3.

While SIC is an approximation, the "derivative discontinuity" is an exact property of the Hohenberg-Kohn density functional $E_v[n]$. This property was discovered by Perdew, Parr, Levy and Balduz (1982),[8] who were motivated in part by the principle of integer preference and in part by the SIC theory which respects this principle. Section 4 reviews the derivative discontinuity, starting from the exact density functional theory for open systems.[8,13] Sections 3 and 4 further examine the striking similarities between SIC and the derivative discontinuity. The relationship of each to the "band-gap problem"[8,13-15] in a semiconductor or insulator is also briefly considered.

Section 5 summarizes the implications of this study for

electronic structure theory in general, and for the construction of size-consistent density functionals in particular.

2. HOHENBERG-KOHN-SHAM THEORY

In order to establish the context and notation for the rest of this article, the spin-density version[16] of the Hohenberg-Kohn-Sham[1,11] theory will be reviewed and analyzed from a size-consistency perspective. The Hamiltonian for N electrons interacting with a spin-dependent external potential $v_\sigma(\vec{r})$ is

$$\hat{H} = \hat{T} + \sum_\sigma \int d^3r \, v_\sigma(\vec{r}) \, \hat{n}_\sigma(\vec{r}) + \hat{V}_{ee} + \ldots, \tag{4}$$

where

$$\hat{T} = \sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2, \tag{5}$$

$$\hat{n}_\sigma(\vec{r}) = \sum_{i=1}^{N} \delta(\vec{r}-\vec{r}_i) \, \delta_{\sigma,\sigma_i}, \tag{6}$$

$$\hat{V}_{ee} = \frac{1}{2} \sum_i \sum_{j \neq i} \frac{1}{|\vec{r}_i - \vec{r}_j|}. \tag{7}$$

The dots in Eq. (4) refer to nonelectronic terms such as the Coulomb repulsion energy of the external charge. The Hohenberg-Kohn theorem[1,16] asserts the existence of a functional

$$E_v[n_\uparrow, n_\downarrow] = G[n_\uparrow, n_\downarrow] + U[n] + \sum_\sigma \int d^3r v_\sigma(\vec{r}) n_\sigma(\vec{r}) \tag{8}$$

of the electron spin densities $n_\sigma(\vec{r}) = \langle \hat{n}_\sigma(\vec{r}) \rangle$, where

$$n(\vec{r}) = \sum_\sigma n_\sigma(\vec{r}), \tag{9}$$

$$U[n] = \frac{1}{2}\int d^3r \int d^3r' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r}'-\vec{r}|} \quad . \tag{10}$$

Minimization of $E_v[n_\uparrow,n_\downarrow]$ subject to the constraint $\int d^3r\, n(\vec{r}) = N$, or equivalently solution of the Euler equation

$$\frac{\delta E_v}{\delta n_\sigma(\vec{r})} = \mu \ , \tag{11}$$

yields the exact ground-state electronic energy and spin-densities.

The universal functional $G[n_\uparrow,n_\downarrow]$ is often split into "noninteracting kinetic" and "exchange-correlation" terms:

$$G[n_\uparrow,n_\downarrow] = T_s[n_\uparrow,n_\downarrow] + E_{xc}[n_\uparrow,n_\downarrow] \ . \tag{12}$$

T_s is the minimum of $<\hat{T}>$ over all wavefunctions (or ensembles) constrained to yield the spin densities $n_\uparrow$ and $n_\downarrow$, while G is the minimum of $<\hat{T} + \hat{V}_{ee}> - U[n]$ from this "constrained search".[2]

In the popular Kohn-Sham method,[11,16] T_s is treated exactly via construction of a noninteracting wavefunction from orbitals that are solutions of the selfconsistent one-electron problem

$$\left[-\frac{1}{2}\nabla^2 + v_\sigma(\vec{r}) + u([n];\vec{r}) + v_{xc}^{\sigma}([n_\uparrow,n_\downarrow];\vec{r})\right]\psi_{\alpha\sigma}(\vec{r}) \tag{13}$$

$$= \varepsilon_{\alpha\sigma}\psi_{\alpha\sigma}(\vec{r}),$$

where the potentials are functional derivatives

$$u([n];\vec{r}) = \delta U/\delta n(\vec{r}) = \int d^3r' \frac{n(\vec{r}')}{|\vec{r}'-\vec{r}|} \quad , \tag{14}$$

$$v_{xc}^{\sigma}([n_\uparrow,n_\downarrow];\vec{r}) = \delta E_{xc}/\delta n_\sigma(\vec{r}) \ . \tag{15}$$

In typical applications, the noninteracting wavefunction is a ground-state Slater determinant, for which

$$T_s = \sum_{\alpha\sigma} f_{\alpha\sigma} \int d^3r \; \psi_{\alpha\sigma}^*(\vec{r})(-\frac{1}{2}\nabla^2) \; \psi_{\alpha\sigma}(\vec{r}) \; , \tag{16}$$

$$n_\sigma(\vec{r}) = \sum_\alpha n_{\alpha\sigma}(\vec{r}) \; , \tag{17}$$

$$n_{\alpha\sigma}(\vec{r}) = f_{\alpha\sigma}(\vec{r})|\psi_{\alpha\sigma}(\vec{r})|^2 \; , \tag{18}$$

$$f_{\alpha\sigma} = \Theta(\mu - \varepsilon_{\alpha\sigma}) \; . \tag{19}$$

The Kohn-Sham orbital energies of Eq. (13) are[17] derivatives of the energy with respect to the occupation number:

$$\varepsilon_{\alpha\sigma} = \partial E/\partial f_{\alpha\sigma}. \tag{20}$$

Within the Kohn-Sham method, only the exchange-correlation energy remains to be approximated. The local spin density (LSD) approximation[11,16] is often applied:

$$E_{xc}^{LSD}[n_\uparrow, n_\downarrow] = \int d^3r \; n\varepsilon_{xc}(n_\uparrow, n_\downarrow) \; , \tag{21}$$

where $\varepsilon_{xc}(n_\uparrow, n_\downarrow)$ is the exchange-correlation energy per particle in an electron gas of uniform spin densities $n_\uparrow$ and $n_\downarrow$. The LSD exchange-correlation potential is

$$\partial[n\varepsilon_{xc}(n_\uparrow, n_\downarrow)]/\partial n_\sigma. \tag{22}$$

The exact exchange-correlation energy may be written[18,19]

$$E_{xc} = \frac{1}{2} \int d^3r \; n(\vec{r}) \int d^3r' \; n_{xc}(\vec{r}, \vec{r}')/|\vec{r}' - \vec{r}| \; , \tag{23}$$

where $n_{xc}(\vec{r}, \vec{r}')$ is the density at $\vec{r}'$ of the "exchange-correlation hole" surrounding an electron at $\vec{r}$. The hole represents a deficit of one electron:

$$\int d^3r' \; n_{xc}(\vec{r},\vec{r}') = -1 \quad . \tag{24}$$

The LSD approximation replaces the exact hole by that of a uniform electron gas; the success of this approximation is explained[19] in part by its satisfaction of the exact sum rule (24). Generalized gradient approximations[9,10] constructed to obey Eq. (24) can be even more accurate[20] than LSD. Fully-nonlocal functionals have also been constructed[21] to satisfy Eq. (24).

The simplicity and practicality of the Kohn-Sham method are due in part to the fact that the potential in Eq. (13) reflects

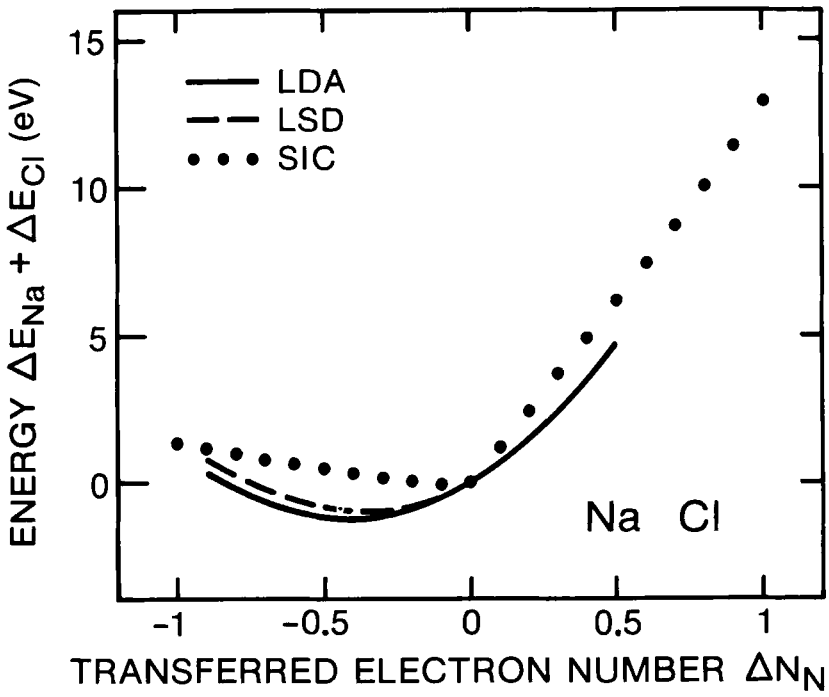


Figure 1. Change in the total energy of a system composed of one sodium atom and one distant chlorine atom, as a function of the number of electrons transferred from chlorine to sodium. LDA: local density approximation. LSD: local spin density approximation. SIC: self-interaction correction to LSD. (From Ref. 13).

the natural symmetries (if any) of the system; this often greatly simplifies the calculation of the orbitals. However, it also implies that the orbitals in a system of separated objects can be delocalized, i.e, one orbital can be distributed over two or more objects (subsystems or open systems).

The noninteracting kinetic energy is separable and extensive, because both size-consistency principles apply even to noninteracting electrons. Consider the application of Eqs. (16)-(19) to a separated subsystem: Those occupied orbitals that are localized within the subsystem present no problem. If the subsystem contains a fragment of an occupied orbital which is shared with other subsystems, then this fragment may be renormalized within the subsystem, and the normalization factor may be absorbed into a "fractional occupation number". After all sums are taken, the subsystem will be described by a collection of localized orbitals, with unit occupation for orbitals with energies less than μ, and possible fractional occupation ($0 \le f \le 1$) for orbitals lying at the chemical potential μ. (A fractional total electron number in an open system is simply the time-average of a fluctuating integer number.)

Local, semi-local, and some nonlocal spin-density functionals for the exchange-correlation energy are properly separable and extensive. However, such approximations typically do not respect the principle of integer preference. For example, consider[5,13] a system composed of a sodium atom and a distant chlorine atom, as described by the local spin density approximation for exchange and correlation. Starting from two neutral atoms, the valence configuration is one Na $3s\uparrow$ electron, one Cl $3p_x\uparrow$ electron, one Cl $3p_y\uparrow$ electron, and three Cl $3p\downarrow$ electrons. Because the energy ε of the Na $3s\uparrow$ orbital is higher than that of the Cl $3p_z\uparrow$ orbital, Eq. (20) shows that the energy of the system will be lowered by transfer of a fraction of an electron from Na $3s\uparrow$ to Cl $3p_z\uparrow$. After about 0.4 electrons have been transferred, the orbital energies are equalized and the total energy is a minimum (Fig. 1).

In a delocalized picture, there is then one electron in a molecular orbital distributed 60% on the Na and 40% on the Cl atom.

Even for a single neutral atom, there are a few examples (Fe, Co) in which the local spin density approximation prefers fractional occupation numbers. In such cases, the occupation numbers should be constrained to integer values,[12,22] even if Eq. (19) is thereby violated.

It should be observed that practical spin-density functionals for the exchange-correlation energy achieve separability at the price of an associated "symmetry dilemma".[23,24]

3. SELF-INTERACTION CORRECTION

An approximate density functional for the exchange-correlation energy may be made exact in the one-electron limit via subtraction of its spurious self-interaction energy. The idea is an old one, and there are many different ways to implement it.[12,25-38] The version of Fermi and Amaldi (1934)[26] depends explicitly upon the total number of electrons in the system, as do some modern implementations;[28,33,35] these radically nonlocal density functionals are not separable. The version of Cortona (1986)[37] is based upon groups of electrons such as atomic subshells, and so is restricted to certain kinds of systems.

Most self-interaction corrections, however, are made on an orbital-by-orbital basis, as in the classical work of Hartree (1927).[25] The total energy is then a functional of the orbitals. Unlike the Hartree-Fock and Kohn-Sham energies, the self-interaction corrected energy is not invariant under unitary transformation of the occupied orbitals. If localized orbitals are energetically preferred over delocalized ones (as in the Hartree approximation), then all three size-consistency principles of section 1 will be respected.

The orbital self-interaction correction (SIC) of Perdew and Zunger (1982)[12,34] will be discussed here. It has been studied

and refined in extensive applications to atoms,[39-55] negative ions,[56-59] molecules,[60,61] metallic clusters,[62] atomic impurities in solids,[63-67] insulators,[68-70] semiconductors,[71,72] and metals.[73-75] SIC is particularly useful whenever the electrons have a natural tendency to localize; it has been applied with interesting results to the Hubbard model[76,77] and to high-temperature superconductors.[77] A "simplified SIC", which crudely simulates the effect of SIC on the energy band structure of a solid, has also been proposed[78] and applied.[78-81] Because a comprehensive study of SIC already exists,[12] it will suffice to sketch the main ideas and a few problems of special interest here.

Begin with an approximate spin-density functional

$$E_{xc}^{approx}[n_\uparrow,n_\downarrow] = E_x^{approx}[n_\uparrow,n_\downarrow] + E_c^{approx}[n_\uparrow,n_\downarrow] \tag{25}$$

for the exchange-correlation energy. In most applications, this is the local spin density (LSD) approximation, but the SIC method is general; recent work by Krieger and Li (1989)[55] suggests that a generalized gradient approximation might be an even better starting point. Now subtract off the spurious self-interaction on an orbital-by-orbital basis:

$$E_x^{SIC} = E_x^{approx}[n_\uparrow, n_\downarrow] - \sum_{\alpha\sigma}\left\{ U[n_{\alpha\sigma}] + E_x^{approx}[n_{\alpha\sigma},0]\right\}, \tag{26}$$

$$E_c^{SIC} = E_c^{approx}[n_\uparrow,n_\downarrow] - \sum_{\alpha\sigma} E_c^{approx}[n_{\alpha\sigma},0]. \tag{27}$$

With this choice, the energy is exact in the one-electron limit. Finally, minimize the energy with respect to variations of the orbitals. If no orthogonality constraint is imposed, the result is a selfconsistent one-electron Schrödinger equation with an orbital-dependent exchange-correlation potential

$$v_{xc}^{\alpha\sigma,SIC}(\vec{r}) = v_{xc}^{\sigma,approx}([n_\uparrow, n_\downarrow];\vec{r}) - u([n_{\alpha\sigma}];\vec{r}) - v_{xc}^{\sigma,approx}([n_{\alpha\sigma},0];\vec{r}). \quad (28)$$

The Hartree approximation[25] is recovered by the choice $E_x^{approx} = E_c^{approx} = 0$. The SIC for LSD exchange was first proposed by Lindgren (1971),[27] and that for correlation by Perdew (1979).[31] Although it is consistent with the Hohenberg-Kohn theorem and with Eq.(20), SIC is not a Kohn-Sham theory.[12]

When applied selfconsistently to atoms, SIC significantly improves the LSD total energy and its separate exchange and correlation components. In the original applications, the orbital densities $n_{\alpha\sigma}(\vec{r})$ of Eq. (18) were spherically averaged, but Harrison (1983)[44,45] has shown that better results are found by retaining their angular dependences.

Now apply SIC to a system composed of two well-separated initially-neutral atoms, one Na and the other Cl. Consider a molecular-orbital variation which gradually transfers an electron from Na $3s\uparrow$ to Cl $3p_z\uparrow$. As Fig. 1 shows, the SIC total energy of the system varies in a nearly-linear way with the occupation f of the $3p_z\uparrow$ orbital. It is easy to understand why: As f varies from 0 to 1, the atomic orbital being filled always "sees" the potential of a 17-electron Cl atom; it does not see its self-generated potential, which has been subtracted out of Eq. (28). Thus the energy of the Cl varies almost as if the extra orbital density were being added to a fixed external potential, i.e., linearly with f. The energy of the Na also varies almost linearly, as does the total energy of the system. If however the electron is transferred the other way, from Cl $3p_z\downarrow$ to Na $3s\downarrow$, the atomic orbital being emptied always sees a 16-electron Cl^+ ion. There is sudden change in effective potential, and thus an abrupt change in the slope of Fig. 1 (derivative discontinuity), when there are exactly 17 electrons on the Cl and 11 on the Na. The total energy minimizes at this cusp, in conformity with the

principle of integer preference. (Note that the LSD and Hartree-Fock approximations, which force all orbitals to see a common potential, cannot achieve this correct limit.)

The SIC analog of Eq. (24) is[12]

$$\int d^3r' \; n_{xc}^{SIC}(\vec{r},\vec{r}') = - \sum_{\alpha\sigma} f_{\alpha\sigma} n_{\alpha\sigma}(\vec{r}) / \; n(\vec{r}). \qquad (29)$$

The right-hand side of Eq. (29) reduces to -1 if no occupation number is fractional. Otherwise, it is greater than -1: In an open subsystem with fractional electron number, part of the exchange-correlation hole is located in distant subsystems with which electrons are shared. Because the LSD hole is always concentrated entirely in the subsystem of interest, the LSD energy is much less realistic for fractional than for integer electron number.

Because the SIC energy E varies almost linearly with occupation $f_{\alpha\sigma}$ of an orbital, it follows immediately from Eq. (20) that the SIC orbital energy $\varepsilon_{\alpha\sigma}$ at full occupation ($f_{\alpha\sigma}=1$) is a good approximation[12] to the negative of the energy needed to remove an electron from orbital $\alpha\sigma$. The LSD orbital energies, on the other hand, are not even close to physical removal energies.

LSD is a convenient method for band-structure calculations in solids, but the fundamental gaps from LSD band-structure calculations for insulators and semiconductors are typically a factor of two smaller than measured gaps. Again, it is not hard to see why:[12] In Kohn-Sham theory, each orbital sees the same potential. In the LSD description of solid neon, an electron just inside a particular unit cell sees the potential of a neutral, 10-electron cell. That is physically correct for an electron in the conduction band, but an electron in the valence band should see only 9 electrons in its unit cell, and not itself. The SIC potential in a localized-orbital description of an insulator has just this correct behavior, and generates the correct band gaps.[12]

Clearly these will be no self-interaction correction to LSD in

a bulk solid if the orbitals are delocalized, as they would be in a Kohn-Sham theory. However, the orbital-dependent SIC potential has enough freedom to break symmetry and generate localized orbitals. These may be atomic-like Wannier orbitals in[68-70] large-gap insulators, and "bond orbitals"[72] in semiconductors.

The mathematical problem to find those orthonormal localized orbitals that minimize the SIC energy has been addressed by Pederson, Heaton and Lin (1984).[60] They start from delocalized "symmetry molecular orbitals" that experience an averaged SIC potential, then perform a unitary transformation to others obeying a localization condition. For an insulator, the symmetry molecular orbitals are Bloch functions with a reasonable band structure and energy gap, while the "localized orbitals" are approximately Wannier functions.

The problem to construct optimum localized orbitals is especially hard for the electron gas of uniform density (jellium).[73-75] Norman (1983)[73] proposed an approximate Wannier function with an orbital density equal to the magnitude of the exchange-hole density. It is clear from his work that the Hartree approximation energetically prefers localized orbitals even for jellium; the Hartree SIC lowers the energy per electron by about $-0.13k_F$ to simulate the correct exchange energy, $-0.239k_F$. More recent work by Pederson, Heaton and Harrison (1989)[75] shows that SIC, applied to LSD exchange, also favors localized orbitals, lowering the energy by at least an additional $-0.005k_F$ to simulate the correlation energy.

When LSD correlation is included, SIC probably[73] favors delocalized orbitals (plane waves) for jellium. If so, then the SIC bulk energy of jellium is exact, but there is still a "false surface energy"[12] which scales like $V^{2/3}$ (whose V is the volume), but depends upon the shape of the system: Consider a spherical shell of jellium with outer radius R and inner radius S, where R, S, and R-S are all macroscopic. When $R/S \to \infty$, there is no self-interaction correction to the surface energy for the cavity of

radius S. But for S=0, there is a positive correction for the solid sphere of radius R. A hint of this effect can be seen in LSD and SIC total energy calculations for small jellium spheres.[62]

4. DERIVATIVE DISCONTINUITY

The discontinuity[8,13,82-84] of the exact Kohn-Sham exchange-correlation potential of Eq. (15) has important consequences[8,13] for the integer preference of open systems,[85] the physical content of the exact Kohn-Sham orbital energies,[47,86,87] and the band gap problem for an insulator or semiconductor.[14,15,88-100] This discontinuity is the doppelganger of the self-interaction correction, accounting for many of the same physical effects within a different theoretical structure. The theory behind the derivative discontinuity has been developed in considerable detail.[13] Here it will suffice to outline that development and discuss a few topics of special interest.

An exact density-functional theory for open systems may be constructed in either of two ways.[13] In the first, the open system is a separated subsystem of a composite with fixed integer electron number. The composite is described by a single wavefunction, and fractional electron number in the open system is generated by variation of this wavefunction. In the second approach, the open system is described via an ensemble or statistical mixture of wavefunctions with different integer electron numbers. The final result for the ground-state energy E of the open system, as a function of its average electron number N, is the same in either approach, since any coherence effects (e.g., infinite-ranged Einstein-Podolsky-Rosen correlations) are energetically irrelevant. Spin-density functionals $T_s[n_\uparrow,n_\downarrow]$ and $E_{xc}[n_\uparrow,n_\downarrow]$ for the open system are constructed by the standard constrained search[2] over wavefunctions or ensembles. E(N) is then the minimum of $E_v[n_\uparrow,n_\downarrow]$ (Eqs. (8) and (12)) over all $n_\uparrow$ and $n_\downarrow$ integrating to N.

The result for E(N) is beautifully simple. If N lies between the integers J and J-1, then the open system is fluctuating beween the ground state of the J-electron system (with probability p=N-J+1) and that of the (J-1) - electron system (with probability 1-p). Therefore its average energy is

$$E(N) = (N - J + 1)\, E(J) + (J - N)\, E(J - 1), \tag{30}$$

and its average density for spin σ is

$$n_\sigma(N,\vec{r}) = (N-J+1)\, n_\sigma(J,\vec{r}) + (J-N)\, n_\sigma(J-1,\vec{r}). \tag{31}$$

Electron numbers other than J and J-1 do not contribute, because of the "convexity" of the energy.[8,13] Thus the energy of an open system varies linearly for N between two integers, much like the SIC energy of Fig. 1, and with many of the same consequences. There is also an exact sum rule[13,83] on the exchange-correlation hole density, which bears an uncanny resemblance to the SIC sum rule of Eq. (29).

Consider the chemical potential $\mu = \partial E/\partial N$ of Eq. (11). By Eq. (20), μ equals $\varepsilon^{HO}(N)$, the highest partially-occcupied Kohn-Sham orbital energy of the N-electron system. But by Eq. (30), this derivative is

$$\mu = \frac{\partial E}{\partial N} = \varepsilon^{HO}(N) = \begin{cases} -I_J & (J-1 < N < J) \\ -I_{J+1} & (J < N < J+1) \end{cases}, \tag{32}$$

where $I_J = E(J-1) - E(J)$ is the "ionization energy of the J-electron system." Eq. (32) displays the physical content of the highest-occupied orbital energy in the exact Kohn-Sham theory. The other Kohn-Sham orbital energies, unlike those of SIC, have little or no physical content.

By Eqs, (8), (11) and (12), the chemical potential of Eq. (32) is

$$\mu = \frac{\delta T_s}{\delta n_\sigma(\vec{r})} + v_\sigma(\vec{r}) + u([n];\vec{r}) + \frac{\delta E_{xc}}{\delta n_\sigma(\vec{r})}. \tag{33}$$

As N rises through the integer J, μ jumps up by the positive constant $I_J - I_{J+1}$. Since the density of Eq. (31) is continuous, as is $u([n];\vec{r})$. it follows that

$$I_J - I_{J+1} = \frac{\delta T_s}{\delta n_\sigma(\vec{r})}\Bigg|_{J-\delta}^{J+\delta} + \frac{\delta E_{xc}}{\delta n_\sigma(\vec{r})}\Bigg|_{J-\delta}^{J+\delta}, \qquad (34)$$

where δ is a positive infinitesimal.

Let the Kohn-Sham orbital energies of the N-electron system be ordered $\varepsilon_1(N) \leq \varepsilon_2(N) \leq \ldots$ Then the argument that led to Eq. (34) may be applied again to the noninteracting system, with the result

$$\varepsilon_{J+1}(J-\delta) - \varepsilon_J(J-\delta) = \frac{\delta T_s}{\delta n_\sigma(\vec{r})}\Bigg|_{J-\delta}^{J+\delta} \qquad (35)$$

Eq. (34) can be rewritten as

$$I_J - I_{J+1} = [\varepsilon_{J+1}(J-\delta) - \varepsilon_J(J-\delta)] + C, \qquad (36)$$

where the constant C is the discontinuity of the exchange-correlation potential (the same for both spins).

Eq. (36) has an important application to the fundamental band gap in an insulator or semiconductor. If J electrons are needed to fill all orbitals up to the top of the valence band, then the left-hand side is the true or measured gap, while the term in square brackets is the prediction of a Kohn-Sham band-structure calculation for the J-electron system. The difference C arises from the jump in the exchange-correlation potential when an extra electron is introduced. Since the separation of G into T_s and E_{xc} is arbitrary, it is clear that such a discontinuity must exist. In a Mott insulator, all of the gap arises therefrom.[8]

The derivative discontinuity implies that the exact Kohn-Sham potential has a radically nonlocal density dependence. Consider again the system composed of one Na and one Cl atom at large mutual separation. Each will be neutral, as the system will be

trapped in the cusp of Fig. 1. The exact Kohn-Sham potential maintains this situation by erecting a "plateau" of positive constant potential[13] in the "domain" of the Cl atom (that region of space in which the density of the Cl dominates that of the Na atom). The height of this plateau is just great enough to make the Cl $3p_z\uparrow$ orbital energy equal to that of Na $3s\uparrow$. Of course the gradient[55,101,102] of the exchange-correlation potential inside this domain is just that of a single Cl atom by itself.

5. SUMMARY OF IMPLICATIONS

The three size-consistency principles discussed in section I are features of nature which should ideally be reflected in electronic-structure theories. The separability and extensivity principles suggest that such theories should not be radically nonlocal; the strongest nonlocality is typically that of the electrostatic energy, which can be treated without approximation. The remarkable successes of local and semi-local density-functional approximations can be explained in part by the fact that they respect these two principles.

The implications of the principle of integer preference depend sensitively upon the kind of theory under consideration. If the kinetic and/or exchange-correlation energies are to be constructed directly from the density, without the intermediation of wavefunctions, then their functional derivatives must have a radically nonlocal dependence upon the density (as discussed in section 4) in order to satisfy this principle. It seems doubtful that such constructions can be found in practice.

However, if the energy is to be constructed entirely from orbitals, and if localized orbitals are energetically preferred over delocalized ones, then all three size-consistency principles can be satisfied without any need for radical nonlocality. The self-interaction corrected local spin density theory has been discussed here as a possible example of such a theory. More

accurate theories of this type could probably be constructed. Unfortunately, the price to be paid for any such theory is the breaking of natural symmetries: Even the electron gas of uniform density becomes a hard (and still unsolved) problem, as discussed in section 3.

For most practical calculations, the traditional Kohn-Sham theory (within or beyond[9,10] LSD) is probably still the best choice. It seems doubtful that practical Kohn-Sham calculations will ever respect the principle of integer preference, but that is often an acceptable price to pay for the simplicity and power of the method.

ACKNOWLEDGMENTS

The ideas and efforts of previous collaborators (especially Mel Levy and Alex Zunger) are gratefully acknowledged, as is the support of the National Science Foundation (DMR88-17866).

REFERENCES

1. P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
2. M. Levy, Proc. Natl. Acad. Sci. USA 76, 6062 (1979).
3. R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. 61, 689 (1989).
4. R. G. Parr and W. Yang, Density Functional Theory of Atoms and Molecules, Oxford U. Press, N. Y. (1989).
5. J. C. Slater, The Self Consistent Field For Molecules and Solids, McGraw-Hill, N. Y. (1974).
6. A. Szabo and N. S. Ostlund, Modern Quantum Chemistry, Macmillan, N. Y. (1982).
7. N. D. Lang, in Theory of the Inhomogeneous Electron Gas, eds. S. Lundqvist and N. H. March, Plenum, N. Y. (1983).
8. J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Jr., Phys. Rev. Lett. 49, 1691 (1982).
9. D. C. Langreth and J. P. Perdew, Phys. Rev. B21. 5469 (1980); D. C. Langreth and M. J. Mehl, Phys. Rev. B28, 1809 (1983).
10. J. P. Perdew and Y. Wang, Phys. Rev. B33, 8800 (1986); J. P. Perdew, ibid., 8822 (1986).
11. W. Kohn and L. J. Sham, Phys. Rev. 140, A 1133 (1965).
12. J. P. Perdew and A. Zunger, Phys. Rev. B23, 5048 (1981).
13. J. P. Perdew in Density Functional Methods in Physics, eds.

R. M. Dreizler and J. da Providencia, Plenum, N. Y. (1985).
14. J. P. Perdew and M. Levy, Phys. Rev. Lett. 51, 1884 (1983).
15. L. J. Sham and M. Schlüter, Phys. Rev. Lett. 51, 1888 (1983).
16. U. von Barth and L. Hedin, J. Phys. C 5, 1629 (1972).
17. J. F. Janak, Phys. Rev. B18, 7165 (1978).
18. D. C. Langreth and J. P. Perdew, Solid State Commun. 17, 1425 (1975); Phys. Rev. B 15, 2884 (1977).
19. O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B13, 4274 (1976).
20. P. Bagno, O. Jepsen, and O. Gunnarsson, Phys. Rev. B40, 1997 (1989) (and references therein).
21. O. Gunnarsson, M. Jonson and B. I. Lundqvist, Phys. Rev. B20, 3136 (1979).
22. S. Dhar, Phys. Rev. A39, 952 (1989).
23. T. Ziegler and A. Rauk, Theoret. Chim. Acta 43, 261 (1977); T. Ziegler, in Local Density Approximations in Quantum Chemistry and Solid State Physics, eds. J. P. Dahl and J. Avery, Plenum, N. Y. (1982).
24. B.I. Dunlap, Phys. Rev. A27, 2217 (1983); B. I. Dunlap, Phys. Rev. A29, 2902 (1984).
25. D. R. Hartree, Proc. Camb. Phil. Soc. 24, 111 (1927).
26. E. Fermi and E. Amaldi, Accad. Ital. Rome 6, 119 (1934).
27. I. Lindgren, Int. J. Quantum Chem. 5, 411 (1971); I. Lindgren and K. Schwarz, Phys. Rev. A5, 542 (1972).
28. M. S. Gopinathan, Phys. Rev. A15, 2135 (1977).
29. H. Stoll, C.M.E. Pavlidou and H. Preuss, Theoret. Chim. Acta 49, 143 (1978).
30. A. Zunger and M. L. Cohen, Phys. Rev. B18, 5449 (1978).
31. J. P. Perdew, Chem. Phys. Lett. 64, 127 (1979).
32. J. F. Dobson and J. H. Rose, J. Phys. C15, L1183 (1982).
33. S. H. Vosko and L. Wilk, J. Phys. B21, 3687 (1983).
34. J. P. Perdew, in Local Density Approximations in Quantum Chemistry and Solid State Physics, eds. J. P. Dahl and J. Avery, Plenum, N. Y. (1984).
35. A. Cedillo, E. Ortiz and J. L. Gázquez, J. Chem. Phys. 85, 7188 (1986).
36. S. B. Trickey, Phys. Rev. Lett. 56, 881 (1986).
37. P. Cortona, Phys. Rev. A34, 769 (1986).
38. S. Manoli and M. A. Whitehead, Phys. Rev. A38, 630 (1988).
39. O. Gunnarsson and R. O. Jones, Solid State Commun. 37, 249 (1981).
40. J. A. Armstrong, S. S. Jha, and K. C. Pandey, Phys. Rev. A23, 2761 (1981).
41. L. A. Cole and B. N. Harmon, J. Appl. Phys. 52, 2149 (1981).
42. L. Wilk and S. H. Vosko, J. Phys. C15, 2139 (1982).
43. J. P. Perdew and L. A. Cole, J. Phys. C15, L905 (1982).
44. J. G. Harrison, J. Chem. Phys. 78, 4562 (1983).
45. J. G. Harrison, J. Chem. Phys. 79, 2265 (1983).
46. J. G. Harrison, R. A. Heaton, and C. C. Lin, J. Phys. B16,

2079 (1983).
47. M. R. Norman and D. D. Koelling, Phys. Rev. B30, 5530 (1984).
48. S. Baroni, J. Chem. Phys. 80, 5703 (1984).
49. V. R. Sawant and D. G. Kanhere, J. Phys. B17, 3003 (1984).
50. M. Levy, R. K. Pathak, J. P. Perdew, and S. Wei, Phys. Rev. A36, 2491 (1987).
51. R. A. Heaton, M. R. Pederson, and C. C. Lin, J. Chem. Phys. 86, 258 (1987).
52. M. R. Pederson and C. C. Lin, J. Chem. Phys. 88, 1807 (1988).
53. K. D. Sen, J. M. Seminario, and P. Politzer, J. Chem. Phys. 90, 4373 (1989).
54. K. A. Jackson and C. C. Lin, Phys. Rev. B39, 1557 (1989).
55. J. B. Krieger and Y. Li, Phys. Rev. A, 39, 6052 (1989).
56. J. P. Perdew, J. H. Rose, and H. B. Shore, J. Phys. B14, L233 (1981).
57. L. A. Cole and J. P. Perdew, Phys. Rev. A25, 1265 (1982).
58. J. G. Harrison, J. Chem. Phys. 84, 1659 (1986).
59. J. G. Harrison, J. Chem. Phys. 86, 2849 (1987).
60. M. R. Pederson, R. A. Heaton, and C. C. Lin, J. Chem. Phys. 80, 1972 (1984).
61. M. R. Pederson, R. A. Heaton, and C. C. Lin, J. Chem. Phys. 82, 2688 (1985).
62. Y. Ishii, S. Ohnishi, and S. Sugano, Phys. Rev. B33, 5271 (1986).
63. A. Zunger, Phys. Rev. B28, 3628 (1983).
64. H. Katayama-Yoshida and A. Zunger, Phys. Rev. Lett. 53, 1256 (1984).
65. L. A. Cole and W. E. Lawrence, Phys. Rev. B30, 4796 (1984).
66. R. A. Heaton, J. G. Harrison, and C. C. Lin, Phys. Rev. B31, 1077 (1985).
67. M. R. Pederson and B. M. Klein, Phys. Rev. B37, 10319 (1988).
68. R. A. Heaton, J. G. Harrison, and C. C. Lin, Phys. Rev. B28, 5992 (1983).
69. R. A. Heaton and C. C. Lin, J. Phys. C17, 1853 (1984).
70. S. C. Erwin and C. C. Lin, J. Phys. C21, 4285 (1988).
71. N. Hamada and S. Ohnishi, Phys. Rev. B34, 9042 (1986).
72. Y. Hatsugai and T. Fujiwara, Phys. Rev. B37, 1280 (1988).
73. M. R. Norman, Phys. Rev. 28, 3585 (1983).
74. J. G. Harrison, Phys. Rev. B35, 987 (1987).
75. M. R. Pederson, R. A. Heaton, and J. G. Harrison, Phys. Rev. B39, 1581 (1989).
76. A. Svane and O. Gunnarsson, Phys. Rev. B37, 9919 (1988).
77. A. Svane and O. Gunnarsson, Europhys. Lett. 7, 171 (1988).
78. M. R. Norman and J. P. Perdew, Phys. Rev. B28, 2135 (1983).
79. M. R. Norman, Phys. Rev. B29, 2956 (1984).
80. J. Hama, Phys. Lett. 105A, 303 (1984).
81. S. J. Mali, G. S. Sohoni, and D. G. Kanhere, Solid State

Commun. 51, 897 (1984).
82. R. G. Parr and L. J. Bartolotti, J. Phys. Chem. 87, 2810 (1983).
83. J. P Perdew and M. Levy in Many-Body Phenomena at Surfaces, eds. D. C. Langreth and H. Suhl, Academic (1984).
84. J. P. Perdew and M. Levy, Phys. Rev. B31, 6264 (1985).
85. J. P. Perdew and J. R. Smith, Surface Sci. 141, L295 (1984).
86. M. Levy, J. P. Perdew, and V. Sahni, Phys. Rev. A30, 2745 (1984).
87. C. O. Almbladh and U. von Barth, in Density Functional Methods in Physics, eds. R. M. Dreizler and J. da Providencia, Plenum, N.Y. (1985).
88. L. J. Sham, Phys. Rev. B32, 3883 (1985).
89. A. E. Carlsson, Phys. Rev. B31, 5178 (1985).
90. J. P. Perdew, Int. J. Quantum Chem. S19, 497 (1986).
91. W. Kohn, Phys. Rev. B33, 4331 (1986).
92. L. Kleinman, Phys. Rev. B33, 7299 (1986).
93. Y.-T. Shen, D. M. Bylander, and L. Kleinman, Phys. Rev. B36, 3465 (1987).
94. M. Rasolt, Phys. Rev. B36, 5041 (1987).
95. O. Gunnarsson and R. S. Jones and S. B. Trickey, Phys. Rev. B36, 3095 (1987).
96. K. Schönhammer and O. Gunnarsson, J. Phys. C20, 3675 (1987).
97. O. Gunnarsson and K. Schönhammer, Phys. Rev. Lett. 56, 1968 (1986); ibid. 60, 1583 (1988).
98. W. Hanke and L. J. Sham, Phys. Rev. B38, 13361 (1988).
99. R. W. Godby and R J. Needs, Phys. Rev. Lett. 62, 1169 (1989).
100. T. K. Ng, Phys. Rev. B39, 1369 (1989).
101. M. K. Harbola and V. Sahni, Phys. Rev. Lett. 62, 489 (1989).
102. Y. Wang, J. P. Perdew, J. A. Chevary, L. D. Macdonald, and S. H. Vosko, unpublished.

DENSITY FUNCTIONAL TREATMENT OF EXCITED STATES

*L. N. Oliveira**

Department of Physics
Santa Barbara, CA 93106

* *Present address: Instituto Física Química São Carlos, Universidade de São Paulo - 13560 São Carlos, SP - Brazil*

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

1. INTRODUCTION

The remarkably successful applications of density functional theory (DFT) to the ground state of atoms, molecules, and solids has fueled interest in extensions to excited states; several alternatives have been proposed. Examples are the Green's function approach, started by Sham and Kohn /1/, and the application of the ground-state formalism to the lowest state of a given symmetry /2/.

Another vein, opened a decade ago by Theophilou /3/, has more recently received increased attention. In order to calculate excited-state energies, he considered the ensemble of M states, equally weighted. For this *equiensemble*, an old extension of the Rayleigh-Ritz principle /4/ generalizes the Hohenberg-Kohn-Sham theory, so that the ensemble averaged energies and densities can (at least formally) be determined. Once these arithmetic averages are known for $M = 1, \ldots, I$, the lowest I individual energies and the corresponding densities can be easily found.

The present chapter is dedicated to the branch of DFT stemming from Theophilou's innovation. It follows the work of Gross, Kohn, and Oliveira, who have extended the Rayleigh-Ritz principle to ensembles of unequally weighted states and the Hohenberg-Kohn-Sham scheme to ensembles of fractionally occupied states /5,6/. These two generalizations are described in Sections 2. and 3., respectively. The extended formalism contains Theophilou's equiensemble approach to the calculation of excited-state energies as a special case. In addition, it provides an alternative procedure (the *fractional-occupation* approach), which relates excitation energies to Kohn-Sham (single-particle) eigenvalues /7/. The latter method is similar to Slater's transition state method /8/, a procedure involving densities computed from fractionally occupied orbitals and excitation energies calculated from single-particle eigenvalues of the Hartree-Fock-Slater equations. Section 3. describes the equiensemble and fractional-occupation schemes and compares the latter to the transition-state approach.

Like the ground-state formalism, for practical applications the ensemble formalism requires an approximation for the exchange-correlation energy functional. Here, unfortunately, the LDA is inadequate. For, since the Mth-excited- and the ground-state energies of a homogeneous electron gas differ infinitesimally, in the LDA the ensemble-averaged and the ground-state exchange-correlation energy functionals are necessarily identical. By contrast, in a small system (*e.g.*, an atom), the ground state differing markedly from the excited states, one expects those two functionals to be very different and the LDA to be unsatisfactory.

This serious problem has led Kohn to develop a quasi-local-density approximation (QLDA) for the equiensemble exchange-correlation functional /9/. That approximation and its relation to the equiensemble and fractional occupation approaches are discussed in Section 4. Section 5. reports an application of both procedures to the excitation spectrum of the He atom /10/. Finally, Section 6. summarizes general conclusions drawn from that test study.

2. RAYLEIGH-RITZ PRINCIPLE FOR EXCITED STATES

The Rayleigh-Ritz principle supporting the Hohenberg-Kohn theorem in Section 3. concerns the spectrum $E_1 \leq E_2 \leq \ldots$ of a Hamiltonian $\hat{H}$ with eigenstates $|j\rangle$, $(j = 1,2,\ldots)$. To state the principle, two definitions are necessary: for given integer m, the set $\mathcal{L}_m$ is defined as the subspace spanned by all the eigenstates $|j\rangle$ with $E_j < E_m$, and $\mathcal{U}_m$ as the subspace spanned by all the eigenstates $|j'\rangle$ with $E_{j'} \leq E_m$. As example, consider a spectrum with $E_1 < E_2 = E_3 = E_4 < E_5 = E_6 < E_7$; for $m = 5$ (or $m = 6$), the following structure defines the two subspaces:

$$\left.\begin{array}{l} E_5 = E_6 \\ \mathcal{L}_m \quad \left\{ \begin{array}{l} E_2 = E_3 = E_4 \\ \\ E_1 \end{array} \right. \end{array}\right\} \mathcal{U}_m.$$

In this notation, here is the Rayleigh-Ritz principle:
(a) *Let $w_1 \geq w_2 \geq \ldots \geq w_M$ be real positive numbers. The following inequality is then satisfied for any set of orthonormal functions $\{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_M\rangle\}$:*

$$w_1\langle\phi_1|H|\phi_1\rangle + \cdots + \langle\phi_M|H|\phi_M\rangle \geq w_1 E_1 + \cdots + w_M E_M. \tag{1}$$

(b) *The equality in (1) holds if and only if for $m = M$, and for all other m with $w_m \neq w_{m+1}$, the subspace $[|\phi_1\rangle, \ldots, |\phi_m\rangle]$, spanned by the trial functions $|\phi_1\rangle, \ldots, |\phi_m\rangle$, lies in $\mathcal{U}_m$ and contains $\mathcal{L}_m$,* i.e.,

$$\mathcal{L}_m \subset [|\phi_1\rangle, \ldots, |\phi_m\rangle] \subset \mathcal{U}_m. \tag{2}$$

This theorem is demonstrated in /5/. Here, to save space, instead of that proof three illustrative particularizations will be presented:

1. $M = 1$ recovers the ground-state Rayleigh-Ritz principle. For non-degenerate ground state, (2) is equivalent to the condition $|\phi_1\rangle = |1\rangle$.
2. Weights $w_1 = \ldots = w_M$ recover the equiensemble principle /4/. In particular, for non-degenerate $|M\rangle$, that condition reduces to $[|\phi_1\rangle, \ldots, |\phi_m\rangle] = [|1\rangle, \ldots, |M\rangle]$.
3. Weights such that $w_1 + \cdots + w_M = 1$ define an operator $\hat{\Gamma} = w_1|1\rangle\langle 1| + \cdots + w_M|M\rangle\langle M|$, with $\mathrm{Tr}\{\hat{\Gamma}\} = 1$. Eq. (1) then implies $\mathrm{Tr}\{\hat{\Gamma}\hat{H}\} \geq E_1$, an inequality established by Löwdin /11/.

3. DENSITY FUNCTIONAL FORMALISM

3.1. The Hohenberg-Kohn Theorem

The variational principle in Section 2. extends the Hohenberg-Kohn theorem to equiensembles as well as to ensembles of fractionally occupied states. A proof of the theorem based on the constrained search

formalism of Levy and Lieb /12,13/ exists /6/, but this chapter, which ignores the v-representability problem, will follow the approach of the seminal article whose 25th anniversary this volume commemorates.

The set $W = \{w_1, \ldots, w_M\}$ defines the ensemble density $\rho^W = w_1\langle 1|\psi^\dagger\psi|1\rangle + \cdots + w_M\langle M|\psi^\dagger\psi|M\rangle$. The Hohenberg-Kohn theorem states that this density determines the external potential $v(r)$ in the Hamiltonian

$$\begin{aligned}\hat{H} &= \int \psi^\dagger(r)\frac{-\nabla^2}{2}\psi(r)\,dr + \int \psi^\dagger(r)v(r)\psi(r)\,dr \\ &\quad + \frac{1}{2}\int \psi^\dagger(r)\psi^\dagger(r')\frac{1}{|r-r'|}\psi(r')\psi(r)\,dr\,dr',\end{aligned} \tag{3}$$

up to within an additive constant. Following the *reductio ad absurdum* reasoning of Hohenberg and Kohn, we assume that the Schrödinger equations for two potentials $v(r)$ and $v'(r)$, differing by more than a constant, have eigenstates $|1\rangle, \ldots, |M\rangle$, and $|1'\rangle, \ldots, |M'\rangle$ that correspond to identical equiensemble densities $\rho^W(r) = \rho'(r)$, respectively /14/.

Referring to the Rayleigh-Ritz principle in Section 2., we now take the eigenstates $|1'\rangle, \ldots, |M'\rangle$ of $\hat{H}' = \hat{T} + \hat{U} + \hat{V}'$ as trial states for the Hamiltonian $\hat{H} = \hat{T} + \hat{U} + \hat{V}$. As shown in /6/, these trial states fail to satisfy condition (2), in part (b) of the Rayleigh-Ritz principle. The *strict* inequality therefore holds in (1), *i.e.*,

$$w_1\langle 1|H|1\rangle + \cdots + w_M\langle M|H|M\rangle < w_1\langle 1'|H|1'\rangle + \cdots + w_M\langle M'|H|M'\rangle. \tag{4}$$

$\hat{H}' - \hat{V}' + \hat{V}$ substituted for H on the right-hand side, we find

$$\begin{aligned}w_1\langle 1|H|1\rangle + \cdots + w_M\langle M|H|M\rangle \;<\; & w_1\langle 1'|H'|1'\rangle + \cdots + w_M\langle M'|H'|M'\rangle \\ & + \int \rho^W(r')[v(r) - v'(r)]\,dr.\end{aligned} \tag{5}$$

Precisely the same reasoning can now be followed with $|1\rangle, \ldots, |M\rangle$ as trial states for the Hamiltonian $\hat{H}'$. We find

$$\begin{aligned}w_1\langle 1'|H'|1'\rangle + \cdots + w_M\langle M'|H'|M'\rangle \;<\; & w_1\langle 1|H|1\rangle + \cdots + w_M\langle M|H|M\rangle \\ & + \int \rho^W(r')[v'(r) - v(r)]\,dr.\end{aligned} \tag{6}$$

Added, Eqs. (5) and (6) produce the inconsistency $0 < 0$, contradicting the assumption that $|1\rangle, \ldots, |M\rangle$ and $|1'\rangle, \ldots, |M'\rangle$ have the same ensemble density. This shows that, apart from an additive constant, the ensemble density $\rho'(r)$ determines uniquely the potential $v'(r)$. For non-

degenerate spectra, it follows that $\rho'(r)$ determines the eigenstates $|1'\rangle,\ldots,|M'\rangle$, so that the ensemble-averaged expectation value

$$E_v^W = w_1\langle 1'|H|1'\rangle + \ldots + w_M\langle M'|H|M'\rangle, \tag{7}$$

which, for $\rho'(r) = \rho^W(r)$, reduces to the ensemble energy

$$\mathcal{E}^W = w_1 E_1 + \ldots + w_M E_M, \tag{8}$$

is a functional $E_v^W[\rho']$ of the ensemble density. Moreover, for any density $\rho'(r) \neq \rho^W(r)$, the Rayleigh-Ritz principle for excited states guarantees the *strict* inequality

$$\mathcal{E}^W = E_v^W[\rho^W] < E_v^W[\rho']. \tag{9}$$

For degenerate spectra, the potential $v'(r)$ does not determine uniquely the eigenstates $|1'\rangle,\ldots,|M'\rangle$. Nonetheless, as /6/ shows, the ensemble-averaged expectation value in Eq. (7) is a (unique) functional of the ensemble density and satisfies the strict inequality in (9).

3.2. The Kohn-Sham Equations

The variational principle (9) leads naturally to Kohn-Sham-like relations determining the equiensemble density $\rho^W(r)$. To generate these single-particle equations, we have to consider a non-interacting electron gas in a potential $v_s[\rho^W](r)$ such that the ensemble density for the non-interacting system equals $\rho^W(r)$. If v_s were known, the single-particle wave-function $\varphi_j(r)$ and energies ε_j for the jth orbital of that system could be easily obtained from the Schrödinger equation

$$[-\frac{\nabla^2}{2} + v_s(r)]\varphi_j(r) = \varepsilon_j\varphi_j(r), \tag{10}$$

and the ensemble density could be subsequently calculated:

$$\rho^W(r) = \sum_{m=1}^{M} w_m \sum_{j=1}^{\infty} f_{jm}|\varphi_j(r)|^2, \tag{11}$$

where the occupation f_{jm} is unity (zero) if the jth orbital is filled (empty) in the mth state of the non-interacting ensemble.

As in the ground-state formalism, the potential v_s is determined variationally. An (approximate) expression for $E_v^W[\rho']$ being therefore necessary, it is convenient to define the ensemble exchange-correlation energy functional $E_{xc}^W[\rho']$. We therefore write

$$E_v^W[\rho'] = T_s[\rho'] + \frac{1}{2}\int \frac{\rho'(r)\rho'(r')}{|r-r'|}\,dr\,dr' + \int \rho'(r)v(r)\,dr + E_{xc}^W[\rho']. \tag{12}$$

Here, $T_s[\rho']$ is the kinetic energy of the non-interacting system:

$$T_s[\rho'] = \sum_{m=1}^{M} w_m \sum_j f_{jm} \int \varphi_j^*(r)(-\frac{\nabla^2}{2})\varphi_j(r)\,dr. \tag{13}$$

To determine this functional of the density $\rho'(r)$, we multiply both sides of Eq. (10) by $\varphi^*(r)$, integrate them over r, and sum them over the occupied orbitals in each state m. Eq. (13) then yields

$$T_s[\rho'] = \sum_{m=1}^{M} w_m \sum_j f_{jm}\varepsilon_j - \int \rho'(r)v_s(r)\,dr, \tag{14}$$

and Eq. (12) becomes

$$E_v^W[\rho'] = \sum_{m=1}^{M} w_m \sum_j f_{jm}\varepsilon_j - \int \rho'(r)[v_s(r)-v(r)]\,dr + \frac{1}{2}\int \frac{\rho'(r)\rho'(r')}{|r-r'|}\,dr\,dr' + E_{xc}^W[\rho']. \tag{15}$$

Applied to this equation, the condition $\delta E_v^W/\delta\rho^W = 0$ now determines the potential $v_s[\rho^W](r)$. From perturbation theory, the first order variation $\delta\varepsilon_j$ of the eigenvalue in (10) due to a change $\delta\rho'$ in the ensemble density is $\delta\varepsilon_j = \int |\varphi_j(r)|^2 \delta v_s(r)\,dr$, and hence

$$\sum_{m=1}^{M} w_m \sum_j f_{jm}\delta\varepsilon_j = \rho^W(r)\delta v_s(r). \tag{16}$$

The variations of the other terms on the right-hand side of Eq. (15) are easily found, leading to

$$v_s[\rho^W](r) = v(r) + \int \frac{\rho^W(r')}{|r-r'|}\,dr' + v_{xc}^W[\rho^W](r) \tag{17}$$

where, as in the ground-state formalism,

$$v_{xc}^{W}[\rho^{W}](r) = \delta E_{xc}^{W}[\rho^{W}]/\delta\rho^{W}(r). \tag{18}$$

Eq. (17) completes the cycle of Kohn-Sham equations that (10) and (11) started. Substitution of (17) for $v_s(r)$ on the right-hand side of Eq. (15) yields a relation similar to the Kohn-Sham expression for the ground-state energy:

$$\begin{aligned}\mathcal{E}^{W} = E_{v}^{W}[\rho^{W}] \;=\; & \sum_{m=1}^{M} w_m \sum_{j} f_{jm}\varepsilon_j - \frac{1}{2}\int \frac{\rho^{W}(r)\rho^{W}(r')}{|r-r'|}\,dr\,dr' \\ & - \int \rho^{W}(r) v_{xc}[\rho^{W}](r)\,dr + E_{xc}^{W}[\rho^{W}].\end{aligned} \tag{19}$$

3.3. Calculation of Excitation Energies

3.3.1. The equiensemble approach

Eq. (19) determines the ensemble-averaged energies $\mathcal{E}^{W}$, linear combinations of the eigenvalues E_j $(j = 1,\ldots,M)$ from which the latter energies can be recovered. As an example, consider the equiensemble, whose identical weights,

$$w_1 = \ldots = w_M = 1/M, \tag{20}$$

suggest a more specific notation for the ensemble-averaged energy, $\mathcal{E}_M$ instead of $\mathcal{E}^{W}$. In this case, the following relation, equivalent to Eq. (8), determines the energy of the g-fold degenerate multiplet encompassing the eigenstates $|M-g+1\rangle,\ldots,|M\rangle$:

$$E_M = \mathcal{E}_M + \frac{M-g}{g}(\mathcal{E}_M - \mathcal{E}_{M-g}). \tag{21}$$

To compute the I lowest energies in a spectrum with degeneracies $g_1,\ldots,g_I$, one then has to calculate $\mathcal{E}_M$ for $M = g_1,\ldots,g_1+\cdots+g_I$. This simple procedure defines the equiensemble approach.

3.3.2. The fractional occupation approach

Inspection of Eq. (21) reveals a shortcoming of the equiensemble method. The expression within parentheses on the right-hand side is a small number obtained by subtracting two large equiensemble energies, associated with two different densities and two different M. The subtraction can therefore introduce large relative errors. This section discusses an alternative procedure (the fractional occupation approach) that avoids this problem.

As Section 4. will show, the currently available expression for the exchange-correlation functional $E_{xc}^{W}[\rho^{W}]$ describes only the equiensemble. This circumstance makes Eq. (21) (involving only equiensemble energies) very convenient and suggests that, to calculate excited state energies, we employ either the weights (20) (the equiensemble approach in Section 3.3.1.) or weights that reduce to the equiensemble in appropriate limits, a possibility the following definition explores:

$$\begin{cases} w_1 = \ldots = w_{M-g} = (1 - wg)/(M - g) \\ w_{M-g+1} = \ldots = w_M = w \end{cases} \qquad (0 \le w \le \frac{1}{M}), \tag{22}$$

where w is a free parameter. For $w = 1/M$, these weights reduce to $1/M$, and we recover Eq. (20). For $w = 0$, the g-fold degenerate multiplet $|M-g+1\rangle,\ldots,|M\rangle$ has zero weight, and the others have weights $w_1 = \ldots = w_{M-g} = 1/(M-g)$, so that we again recover Eq. (20), with $M \to M - g$. Eq. (22) hence interpolates between equiensembles. The added flexibility supplied by the parameter w opens a new avenue leading to excitation energies, as the following discussion shows.

With the weights (22) substituted on the right-hand side, Eq. (8) can be differentiated with respect to w. We find

$$\frac{d\mathcal{E}^{W}}{dw} = gE_M - \frac{g}{M-g}\sum_{m=1}^{M-g} E_m \tag{23}$$

and, after substituting Eq. (21) for E_M on the right-hand side,

$$\frac{d\mathcal{E}^{W}}{dw} = M(\mathcal{E}_M - \mathcal{E}_{M-g}). \tag{24}$$

Now, $\mathcal{E}^{W}$ is given by Eq. (19), which with the present weights reads

$$E_v^{W}[\rho^{W}] = \frac{1-wg}{M-g}\sum_{m=1}^{M-g}\sum_j f_{jm}\varepsilon_j + w\sum_{m=M-g+1}^{M}\sum_j f_{jm}\varepsilon_j + E_{xc}^{W}[\rho^{W}]$$

$$-\int \frac{\rho^W(r)\rho^W(r')}{2|r-r'|}\,dr\,dr' - \int \rho^W(r) v_{xc}[\rho^W](r)\,dr.$$

By differentiating this equality with respect to w we can therefore calculate the left-hand side of Eq. (24). Although the ensemble density $\rho^W(r)$ (produced by the Kohn-Sham equations) depends on w, this derivative is very easily computed, because the variational property of $E_{xc}^W[\rho^W]$ ensures that $dE_v^W[\rho^W]/dw = \partial E_v^W[\rho^W]/\partial w$. It follows that

$$\frac{dE_v^W[\rho^W]}{dw} = \sum_{m=M-g+1}^{M}\sum_j f_{jm}\varepsilon_j - \frac{g}{M-g}\sum_{m=1}^{M-g}\sum_j f_{jm}\varepsilon_j + \left.\frac{\partial E_{xc}^W[\rho^W]}{\partial w}\right|_{\rho^W(r)}, \tag{25}$$

and from Eq. (24) that

$$M(\mathcal{E}_M - \mathcal{E}_{M-g}) = \sum_{m=M-g+1}^{M}\sum_j f_{jm}\varepsilon_j - \frac{g}{M-g}\sum_{m=1}^{M-g}\sum_j f_{jm}\varepsilon_j + \left.\frac{\partial E_{xc}^W[\rho^W]}{\partial w}\right|_{\rho^W(r)} \tag{26}$$

To compute the energy difference on the right-hand side of (21) it is hence sufficient to solve the Kohn-Sham equations for a single w in the interval $0 \le w \le 1/M$, so that (26) avoids the numerical problem in the equiensemble approach. Practical applications of this equality of course require an approximation for $\partial E_{xc}^W[\rho^W]/\partial w$. For $w = 1/2M$, Section 4. presents a simple expression for this derivative.

4. THE QUASI-LOCAL DENSITY APPROXIMATION

4.1. The Equiensemble Exchange-Correlation Functionals

This section discusses the QLDA for the ensemble-averaged exchange-correlation energy functional /9/. This approximation replaces ensemble averages (for which, as Section 1. pointed out, the LDA is inadequate) by thermal averages (for which the LDA is appropriate). It is

therefore expected to become accurate for $M \to \infty$, the limit in which the thermal ensemble and the equiensemble become identical.

In the Kohn-Sham scheme of Section 3.2., *two* ensembles of identical multiplicity M are involved, one of interacting and the other of non-interacting particles. With them we will associate two canonical ensembles, at temperatures θ and θ_s, respectively. Each equiensemble has entropy $k_B \ln M$, where k_B is Boltzmann's constant. Therefore, the following LDA fixes the two temperatures:

$$\int \sigma^{\theta}[\rho^W](r)\,dr = \int \sigma_s^{\theta_s}[\rho^W](r)\,dr = k_B \ln M. \tag{27}$$

Here and henceforth, the symbols $[\rho](r)$ denote the LDA, indicating the preceding quantity [in this case σ^{θ} ($\sigma_s^{\theta_s}$), the entropy per unit volume of interacting (non-interacting) electrons at temperature θ (θ_s)] for a homogeneous system with the local density $\rho(r)$.

Eq. (12) expresses the exchange-correlation energy functional in terms of the equiensemble energy $E_v^W[\rho']$ and of the kinetic energy $T_s^W[\rho']$. For large M, the former functional is approximately equal to the canonical-ensemble energy functional $E_v^{\theta}[\rho']$, in turn related to the canonical-ensemble exchange-correlation energy functional $E_{xc}^{\theta}[\rho']$ by the thermal analogue of (12) /15/:

$$E_v^{\theta}[\rho'] = T_s^{\theta}[\rho'] + \int \frac{\rho'(r)\rho'(r')}{2|r-r'|}\,dr\,dr' + \int \rho'(r)v(r)\,dr + E_{xc}^{\theta}[\rho']. \tag{28}$$

The right-hand side of this equation substituted for $E_v^W[\rho']$ and the canonical-ensemble kinetic-energy functional at temperature θ_s substituted for $T_s^W[\rho']$, Eq. (12) becomes

$$E_{xc}^W[\rho'] = T_s^{\theta}[\rho'] + E_{xc}^{\theta}[\rho'] - T_s^{\theta_s}[\rho'], \tag{29}$$

thus relating the equiensemble exchange-correlation energy functional to canonical-ensemble functionals, for which the LDA is appropriate. For systems of slowly varying density, therefore,

$$E_{xc}^W[\rho'] \approx \int \{t_s^{\theta}[\rho'](r) + e_{xc}^{\theta}[\rho'](r) - t_s^{\theta_s}[\rho'](r)\}\,dr, \tag{30}$$

where t_s^{θ} (e_{xc}^{θ}) denotes the kinetic (exchange-correlation) energy per unit volume for a homogeneous non-interacting (interacting) system at temperature θ.

Eqs. (18) and (30) determine the equiensemble exchange-correlation potential. Since W represents the equiensemble, the derivative on the right-hand side of (18) must be taken at *fixed entropy*. With the thermodynamical identity $\mu = \partial E(\rho,S)/\partial\rho|_S$ (relating internal energy E, entropy S, and chemical potential μ), simple manipulations lead to

$$v_{xc}^W[\rho'](r) \approx \mu_s^\theta[\rho'](r) + \mu_{xc}^\theta[\rho'](r) - \mu_s^{\theta_s}[\rho'](r), \tag{31}$$

where μ_s^θ (μ_{xc}^θ) denotes the chemical potential (the exchange-correlation chemical potential) for homogeneous non-interacting (interacting) electrons at temperature θ.

This approximation for the exchange-correlation potential and Eq. (30) for the exchange-correlation energy functional complement the Kohn-Sham equations for the equiensemble approach. The two expressions, weakly non-local because the temperatures θ and θ_s are non-local functionals of the density [see (27)], are made practical by satisfactorily accurate approximations for the various thermal properties of the homogeneous interacting and non-interacting electron gases appearing on the right-hand sides /16/.

For $M = 1$ ($\theta = \theta_s = 0$), Eqs. (30) and (31) reduce to the LDA for the ground-state exchange-correlation energy functional and potential, respectively. The QLDA is therefore accurate both for $M \to 1$ and for $M \to \infty$; it should be satisfactory even for moderately large M. The numerical results in Section 5. support this conclusion.

4.2. Application to the Fractional Occupation Method

4.2.1. Approximate form for $\partial E_{xc}^W[\rho^W]/\partial\rho^W(r)$

As we have just seen, the QLDA provides an expression for the equiensemble exchange-correlation functional, hence making practical the equiensemble approach. In addition, as the present section shows, it produces an approximation for the last term on the right-hand side of Eq. (26), hence making practical the fractional occupation approach.

Section 3.3.2. has pointed out that, for $w = 0$ and $1/M$, the weights w_m in Eq. (22) reduce to the equiensembles with multiplicities $M - g$ and M, respectively. To these the QLDA applies, suggesting the following expression for the partial derivative in (26):

$$\left.\frac{\partial E_{xc}^{W}[\rho^{W}]}{\partial w}\right|_{\rho^{W}} = M\{E_{xc}^{w=1/M}[\rho^{w=1/2M}] - E_{xc}^{w=0}[\rho^{w=1/2M}]\} + O(1/M^2). \quad (32)$$

Although the right-hand side involves the energy functionals for $w = 0$ and $1/M$, *i.e.*, $E_{xc}^{M-g}[\rho']$ and $E_{xc}^{M}[\rho']$, respectively, both functionals are evaluated at the density $\rho^{w=1/2M}(r)$, *i.e.*, at the ensemble density found by solving the Kohn-Sham equations in Section 3.2. with $w = 1/2M$. This weight eliminates the term of $O(1/M)$ from (32).

To compute $\rho^{w=1/2M}$ we also need the exchange-correlation potential $v_{xc}^{w=1/2M}$. In analogy with Eq. (32), we can write

$$v_{xc}^{w=1/2}[\rho^{W}](r) = \frac{v_{xc}^{w=0}[\rho^{W}](r) + v_{xc}^{w=1/M}[\rho^{W}](r)}{2} + O(1/M^2). \quad (33)$$

With $v_{xc}^{w=0}$ and $v_{xc}^{w=1/M}$ given by (31), Eqs. (32) and (33) turn the fractional occupation approach into a practical computational procedure. An illustration will be presented in Section 5.. Before that, however, the next section discusses an important alternative.

4.2.2. Comparison with the transition state approach

Consider an N-electron interacting system with non-degenerate ground state and g-fold degenerate first excited state. Applied to the ensemble multiplicity $M = 1 + g$, (26) gives its first excitation energy:

$$\overline{E}_2 - E_1 = \varepsilon_{N+1} - \varepsilon_N + \frac{1}{g}\left.\frac{\partial E_{xc}^{W}[\rho^{W}]}{\partial w}\right|_{\rho^{W}(r)}, \quad (34)$$

where $\overline{E}_2 = E_2 = \ldots = E_M$ is the second-multiplet energy. ε_N (ε_{N+1}) is the Nth ($N+1$th) Kohn-Sham eigenvalue, *i.e.*, the last occupied (first vacant) level in the non-interacting ground state. Both correspond to the ensemble density $\rho^{W} = (1 - wg)\rho_1 + wg\overline{\rho}_2$, where ρ_1 ($\overline{\rho}_2$) is the ground-state (average first-excited-multiplet) density. Compare this equation with the transition state formula /8/:

$$\overline{E}_2 - E_1 = \varepsilon^S_{N+1} - \varepsilon^S_N, \tag{35}$$

where ε^S_N denotes the Nth eigenvalue of the Hartree-Fock-Slater equations for the density $\rho^S = (\rho_1 + \overline{\rho}_2)/2$. Three distinctions separate Eqs. (34) and (35): (1) the Hartree-Fock-Slater equations take only ground-state exchange into account, whereas the Kohn-Sham like equations involve the ensemble averaged exchange-correlation energy functional; (2) for $w \neq 1/2g$, the densities ρ^S and ρ^W are different, and (3) the last term on the right-hand side of Eq. (34) is missing from Eq. (35). Nevertheless, for $w = 1/2M \equiv 1/2(g+1)$, Eq. (32) provides an estimate of that term; as it turns out, its contribution to Eq. (34) is small /10/. Thus, for highly degenerate first excited state (*i.e.*, for $M \approx g$), Eqs. (34) and (35) yield approximately equal results.

From this, one might conclude that the fractional occupation method is a refinement of the transition state method. This inference cannot be generalized, however: to calculate the Ith energy in the spectrum, the transition state approach prescribes $\rho^S = (\rho_1 + \overline{\rho}_I)/2$, while ρ^W is the ensemble average of the densities $\rho_1, \overline{\rho}_2, \ldots, \overline{\rho}_I$. Moreover, for $I > 2$, (26) is very different from the equivalent of (35) /8/. For energies higher than $\overline{E}_2$, therefore, the two methods can be compared only on the basis of the numerical results they produce. This considered, a more conclusive comparison will be deferred to Section 5.2..

5. NUMERICAL RESULTS FOR THE He ATOM

5.1. Equiensemble Energies

The He atom, a small system with widely-studied excitation spectrum and negligible spin-orbit interaction /18/, is a convenient testground for the approximations in Section 4.. The numerical computation /10/ of its equiensemble energies by the two methods in Section 6 will be reported in this section. To recapitulate, in the equiensemble pro-

cedure, Eq. (19) yields the ensemble-averaged energies. In the fractional occupation approach, Eq. (26) determines equiensemble-energy differences; from these, the equiensemble energies measured from a reference level, such as the ground-state energy, can be computed.

Those expressions make no reference to electron spin; ensemble-density-functional theory has so far been limited to spin-averaged densities. This restricts the following discussion to Helium densities and energies averaged over the *para* ($S = 0$) and *ortho* ($S = 1$) configurations of each excited state, weights one and three, respectively.

The energies are then labeled by principal quantum number n and angular momentum L. Each multiplet nL has degeneracy $g = 4(2L + 1)$. In increasing-energy order, the degeneracies $g = 1$ [the 1S (ground) state], $g = 4$ (2S), $g = 12$ (2P), ..., $g = 4$ (5S) form the ensemble multiplicities $M = 1, 5, 17, \ldots, 121$. These and the ensemble-averaged experimental energies (reckoned from the experimental ionization threshold, $E_1^{+\,EXPT} = -4\,\mathrm{Ry}$) appear in Table I. The equiensemble energies [measured from the ground-state energy ($E_1^{+\,LDA} = -3.725\,\mathrm{Ry}$) for the ionized He atom, calculated in the LDA /19/] are also tabulated.

Subtracting E_1^+ from each set of equiensemble energies eliminates from Table I the unphysical self-interactions contained in the energies calculated with the LDA /17/ (for example, most of the discrepancy between $E_1^{+\,EXPT}$ and $E_1^{+\,LDA}$ is self-interaction) and, by extension, in those calculated with the QLDA. This focuses attention on specific properties of the latter approximation.

The agreement between the calculated and experimental equiensemble energies goes from good to excellent as M increases. This is of course expected, the QLDA being accurate for large M. Although M-dependent, the deviation between the calculated and the experimental $\mathcal{E}_M$ follows a pattern: for growing M, it is significantly reduced whenever a new angular momentum is introduced. The deviation, nearly 10% for the (two) equiensembles involving only S states, decreases to 5% for those involving S and P states, to 2% for S, P, and D states, and to 1% for S, P, D, and F states. This seems to be due to the LDA embedded in the QLDA, for an ensemble rich in angular momenta should be well represented by a homogeneous electron gas.

Table I: Equiensemble energies[a] and energy differences in the equiensemble (EQ) and fractional occupation (FO) approaches, compared with the experimental (EXPT) values.

M	Multiplet[b]	$\mathcal{E}_M^{EXPT}$	$\mathcal{E}_M^{EQ}$	$\Delta\mathcal{E}_M^{EQ}$	$\Delta\mathcal{E}_M^{FO}$	$\Delta\mathcal{E}_M^{EXPT}$
1	1S	-1.807	-1.947(8%)			
5	2S	-0.630	-0.727(15%)	1.220	1.240	1.177
17	2P	-0.370	-0.386(4%)	0.341	0.346	0.260
21	3S	-0.325	-0.345(6%)	0.041	0.041	0.045
33	3P	-0.249	-0.265(6%)	0.081	0.083	0.077
53	3D	-0.197	-0.200(2%)	0.065	0.067	0.052
57	4S	-0.188	-0.192(2%)	0.008	0.008	0.009
69	4P	-0.166	-0.171(3%)	0.020	0.020	0.022
89	4D	-0.143	-0.146(2%)	0.025	0.024	0.023
117	4F	-0.124	-0.123(1%)	0.024	0.023	0.019
121	5S	-0.121	-0.120(1%)	0.002	0.002	0.003

[a] Energies (Ry) measured from the ionization threshold. Percentages show deviations from experimental values.
[b] Highest multiplet in equiensemble.

This non-uniform convergence of the calculated equiensemble energies to the experimental values affects perniciously the equiensemble-energy differences; Eq. (21) shows that the calculated excited-state energies will be likewise degraded. In light of the introductory argument in Section 3.3.2., one might expect Eq. (26) to yield more accurate equiensemble-energy differences. Table I indicates that it does not, apparently because the approximation (32) for $\partial E_{xc}^{W}[\rho^{W}]/\partial\rho^{W}(r)$ inherits the limitations of the QLDA. In particular, the last two terms in that equation involve ensemble multiplicities M and $M-g$, respectively, so that the fluctuating accuracy of the QLDA defiles the equiensemble-energy differences calculated with the fractional occupation method just as it corrodes precision in the equiensemble method. From this standpoint, the agreement between the equiensemble energy differences calculated by the two procedures is disappointing.

Table II: Eleven lowest energies[a] in the spectrum of the He atom, calculated by the equiensemble (EQ), fractional occupation (FO), and transition state (TS) methods, compared with the experimental values[b] (EXPT).

Multiplet	I	ΔE_I^{EXPT}	ΔE_I^{EQ}	ΔE_I^{FO}	ΔE_I^{TS}
1S	1	1.807	1.947(8%)	1.972(9%)	
2S	2	0.336	0.422(26%)	0.422(26%)	0.263(22%)
2P	3	0.262	0.244(7%)	0.243(7%)	0.171(35%)
3S	4	0.134	0.171(27%)	0.174(29%)	0.096(28%)
3P	5	0.115	0.123(7%)	0.118(3%)	0.076(34%)
3D	6	0.111	0.093(16%)	0.086(23%)	0.067(40%)
4S	7	0.072	0.086(19%)	0.079(10%)	0.051(29%)
4P	8	0.064	0.075(17%)	0.073(14%)	0.040(38%)
4D	9	0.063	0.059(6%)	0.060(5%)	0.037(41%)
4F	10	0.063	0.048(24%)	0.047(25%)	0.034(46%)
5S	11	0.045	0.047(4%)	0.045[c]	0.032(30%)

[a] All energies (Ry) subtracted from the ionization threshold, *i.e.*, $\Delta E_I \equiv \overline{E}_I - E_1^+$.

[b] Percentages show deviations from experimental energies.

[c] Reference energy for fractional occupation approach.

5.2. Excitation Spectrum

Table II displays the multiplet energies $\overline{E}_I$, $(I = 1,\ldots,11)$, derived from the equiensemble energies and energy differences in Table I. All levels have been subtracted from the ionization threshold. Since the fractional occupation approach produces only equiensemble energy differences, to compute the energy on the left-hand side of Eq. (21) it is necessary to measure them from a reference level. The calculation becoming more accurate for large M, it is convenient to reckon the equiensemble energies from the the largest experimental energy in the spectrum, *i.e.*, in Table II, from E_{11}^{EXPT}. In practice, this makes $E_{11}^{FO} = E_{11}^{EXPT}$.

As already discussed in Section 5.1., the mutual agreement between the energies produced by the two approaches is excellent. The inferior agreement between the experimental and calculated equiensemble energy differences $\Delta\mathcal{E}_M$ accounts for the deviations ($\approx 20\%$) between the experimental excited state energies and those calculated by either method.

Table II also compares the two density-functional approaches with Slater's transition-state method. By substituting the ground-state energy functional for the ensemble energy functional, the latter avoids the numerical problem just discussed and produces more uniform accuracy. In spite of this, as the table indicates, the equiensemble and fractional occupation approaches are more reliable.

6. CONCLUSIONS

The formally exact expressions in Section 3., complemented by the approximations in Section 4., define two ensemble approaches to the calculation of excitation energies. Both schemes, the equiensemble and the fractional occupation procedures, have been applied to the He atom (Section 5.).

For large ensemble multiplicity, the QLDA is very good. Its accuracy is, nevertheless, M-dependent. As a consequence, in the equiensemble approach the equiensemble-energy differences contain large percent errors; through Eq. (21) these are subsequently transmitted to the calculated excited state energies.

In principle, the fractional occupation procedure can avoid this problem. To be practical, however, that approach depends on approximations for the derivative $\partial E_{xc}^W[\rho^W(r)]/\partial w$. The illustrative form proposed in Section 5.2. involves two approximations: (1) neglecting the terms of $\mathcal{O}(1/M^2)$ in Eqs. (32) and (33) and (2) substituting the QLDA form (30) for $E_{xc}^W[\rho^W]$. As the numerical results indicate, the former approximation is excellent. The latter, however, suffers from

the limitation of the QLDA, which makes the fractional occupation and equiensemble approaches essentially equivalent.

In summary, the two procedures produce fairly accurate excited state energies (typical error of 20%). Both methods admit improvements, provided only that (1) a uniformly accurate approximation, to replace the QLDA, or (2) a better approximation for $\partial E_{xc}^{W}[\rho^{W}]/\partial w$ be developed. Unfortunately, the uniformity in the former alternative imposes a very unwieldy constraint. The second research line seems more promising.

ACKNOWLEDGMENTS

Many ideas in this chapter are Hardy Gross's and Walter Kohn's brainchildren. Most others emerged in discussions among the three of us. I am honored to have collaborated with them and deeply indebted to their incentive and hospitality. The NSF (contract DMR-87-03434) and the CNPq (Brazil) supported this work.

REFERENCES

1. L. J. Sham and W. Kohn, Phys. Rev. **145**, 561 (1966).
2. O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B **13**, 4274 (1976), H. Englisch and R. Englisch, Physica A **121**, 253 (1983).
3. A. K. Theophilou, J. Phys. C **12**, 5419 (1978); N. Hadjisavvas and A. K. Theophilou, Phys. Rev. A **32**, 720 (1985).
4. R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1965), Vol. I, p. 459.
5. E. K. U. Gross, L. N. Oliveira, and W. Kohn, Phys. Rev. A **37**, 2805 (1988).

6. E. K. U. Gross, L. N. Oliveira, and W. Kohn, Phys. Rev. A **37**, 2809 (1988).
7. J. Katriel, J. Phys. C **13**, L375 (1980), has identified the equiensemble with the M-particle ground-state of a "super-Hamiltonian". With this identification, the ensemble of fractionally occupied states corresponds to an ensemble of open systems [see J. P. Perdew, Int. J. Quantum Chem. Symp. **19**, 497 (1986) and references cited].
8. J. C. Slater, *The Self-Consistent Field for Molecules and Solids* (McGraw-Hill, New York, 1974), Vol. IV, p. 51.
9. W. Kohn, Phys. Rev. A **34**, 737 (1986).
10. L. N. Oliveira, E. K. U. Gross, and W. Kohn, Phys. Rev. A **37**, 2821 (1988).
11. P. O. Löwdin in *Density Matrices and Density Functionals*, edited by R. Erdahl and V. H. Smith, Jr., (D. Reidel, 1987), p. 21.
12. M. Levy, Phys. Rev. A **26**, 1200 (1982).
13. E. H. Lieb, in *Physics as Natural Philosophy*, edited by A. Shimony and H. Feshbach (MIT, Cambridge, 1982), p. 111.
14. To avoid the cumbersome notation $\rho^{v^W}(r)$, the ensemble density associated with the potential $v^l(r)$ is denoted $\rho^l(r)$.
15. For a review, see W. Kohn and P. Vashishta, in *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N. H. March (Plenum, New York, 1983), p. 79.
16. M. W. C. Dharma-wardana and R. Taylor, J. Phys. C **14**, 629 (1981); F. Perrot and M. W. C. Dharma-wardana, Phys. Rev. A **30**, 2619 (1984); S. Tanaka, S. Mitake, and S. Ichimaru, Phys. Rev. A **32**, 1896 (1985); R. G. Dandrea, N. W. Ashcroft, and A. E. Carlsson, Phys. Rev. B **34**, 2097 (1986).
17. J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
18. W. C. Martin, J. Phys. Chem. Ref. Data, **2**, 257 (1973).
19. S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).

SELF-ENERGY APPROACH TO QUASIPARTICLE ENERGIES USING A DENSITY FUNCTIONAL TREATMENT OF DIELECTRIC SCREENING

Mark S. Hybertsen

AT&T Bell Laboratories
Murray Hill, New Jersey

Steven G. Louie

Department of Physics
University of California
and Materials and Chemical Sciences Division
Lawrence Berkeley Laboratory
Berkeley, California

1. INTRODUCTION

The proper treatment of the particle like excitations of interacting electrons has been a problem of long standing difficulty and interest [1]. The concept of a quasiparticle description of the excited states has proven to be appropriate for describing a wide range of experimental probes including transport, photoemission and optical response. Nonetheless, actual calculation of the quasiparticle energies requires adequate treatment of the dynamical correlations amongst the electrons. This has proven difficult even for the simplest model of a uniform electron gas let alone the inhomogeneous electron system that results in real condensed matter applications due to the electron-ion interaction

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

and chemical bonding [1].

For treatment of the ground state energy, the approach based on the density functional theory (DFT) [2] has been widely successful, largely in conjunction with the local density approximation (LDA) for the exchange-correlation contribution to the energy [3]. As discussed in various articles in this volume, a broad range of solid state systems and ground state properties have been calculated in excellent agreement with experiment. However the eigenvalues that result from the Kohn-Sham implementation of the density functional theory do not yield the quasiparticle energies [4]. Nonetheless, they are often used in this manner, for lack of an alternative parameter free band structure technique. As a consequence, a broad base of experience has been accumulated which demonstrates the deviation of the LDA eigenvalues from true quasiparticle energies [3]. A simple example is the electron gas model where symmetry requires the true exchange-correlation potential to be constant implying that the DFT eigenvalues are identical to the free electron energies up to a constant shift. However, it is well known that the electron-electron interaction yields a renormalized effective electron mass [1]. For the case of alkali metals, the LDA band width is found to be considerably too large in comparison to recent photoemission measurements [5]. Finally, a dramatic example is the underestimate of semiconductor and insulator band gaps by 30-100% in comparison to experiment. Within the DFT framework, the error for the minimum gap has been shown to be largely due to a discontinuity in the true DFT exchange-correlation potential [6-8].

It is clear that a direct approach for obtaining quasiparticle energies is required. In this paper, we give a brief review of a self energy approach which explicitly treats the many-electron correlations that are essential to the quasiparticle energies. This is *not* an "improved" density functional. Rather, we provide a direct treatment of the many-electron problem. However, the LDA is used as a tool to provide the necessary input to the electron self energy operator, in particular the proper description of the dynamically screened Coulomb interaction. Nonetheless, the relationship of the quasiparticle energies to the DFT or

LDA eigenvalues will be pointed out where interesting. The present approach has been applied to a broad range of materials systems with excellent results in comparison to experiment including bulk semiconductors and insulators [9,10], simple metals [11], semiconductor interfaces [12] and superlattices [13], and semiconductor surfaces [14]. These results not only allow elucidation of the dynamical correlations involved, but also demonstrate that a predictive theory is possible with applications to real materials systems. In particular, the LDA may be used to establish ground state properties, e.g. the atomic configuration, of novel materials. The present self energy approach is a complementary tool for predicting spectroscopic properties.

2. THE SELF ENERGY APPROACH

The quasiparticles in a solid are described by a Schroedinger type equation [1]:

$$(T + V_{ext} + V_H)\psi_{n\mathbf{k}}(\mathbf{r}) + \int d\mathbf{r}'\Sigma(\mathbf{r}, \mathbf{r}'; E_{n\mathbf{k}})\psi_{n\mathbf{k}}(\mathbf{r}') = E_{n\mathbf{k}}\psi_{n\mathbf{k}}(\mathbf{r}), \quad (1)$$

including the kinetic energy, the external potential due to the ion cores, the average electrostatic (Hartree) potential, and the non-local, energy dependent and in general complex self energy operator. The self energy operator Σ describes the exchange and correlations among the electrons. The real part of the eigenvalue gives the energy of the quasiparticle while the imaginary part corresponds to the lifetime.

The physically important approximation is for the self energy operator. Here, following Hedin [15], the self energy is developed in a perturbation series in terms of the dynamically screened Coulomb interaction, W, and the full Green's function, G. This is expected to provide better convergence properties than an expansion in the bare Coulomb interaction. The first term only is retained, the so-called GW approximation [1,15]:

$$\Sigma(\mathbf{r}, \mathbf{r}'; E) = i\int \frac{dE'}{2\pi} e^{-i\delta E'} G(\mathbf{r}, \mathbf{r}'; E-E')W(\mathbf{r}, \mathbf{r}'; E'), \quad (2)$$

where $\delta=0^+$. In this approximation, vertex corrections to the self energy are neglected. The W is fully described by the dielectric matrix ϵ^{-1} which screens v, the bare Coulomb interaction:

$$W(\mathbf{r}, \mathbf{r}'; E) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; E) v(\mathbf{r}''-\mathbf{r}'). \tag{3}$$

For the Green's function, a quasiparticle approximation is used:

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_{n,\mathbf{k}} \frac{\psi_{n\mathbf{k}}(\mathbf{r})\psi^*_{n\mathbf{k}}(\mathbf{r}')}{E - E_{n\mathbf{k}} - i\delta_{n\mathbf{k}}}. \tag{4}$$

Here $\psi_{n\mathbf{k}}$, $E_{n\mathbf{k}}$ are quasiparticle wavefunctions and energies implicitly defined by Eq. (1), $\delta_{n\mathbf{k}}=0^+$ for occupied states and $\delta_{n\mathbf{k}}=0^-$ for empty states.

The full dynamical dielectric matrix is required. This represents a major difficulty for evaluating the self energy operator for a real material such as a semiconductor. Due to the charge inhomogeneity, the so-called local fields in the screening response are physically very important. These local fields yield the variation in the screened Coulomb interaction around an added electron as it moves from regions of high density, e.g. bonds, to regions of low density, e.g. interstitial [16]. The local fields are described by the off-diagonal elements of the dielectric matrix in a reciprocal space representation. We obtain the dynamical response in two steps. First, the static dielectric matrix is calculated as a ground state property from the LDA [16]. Second, the dielectric matrix is extended to finite frequency using a generalized plasmon pole model [9] which represents each component $\mathbf{q}$, $\mathbf{G}$, $\mathbf{G}'$ of the imaginary part of the dielectric matrix by a delta function. No adjustable parameters enter. A variation on this approach has been introduced [17] with similar results.

Some progress has been made in finding a model for the screening which contains the essential role of local fields. Since the plasmon pole model is found to be adequate, attention can be focussed on the static response function. The important part of the local fields for the self energy problem is the variation in the local screening potential near the site of the added electron as the site is varied. This is illustrated by comparing Fig. 1a to Fig. 2a where the static screening potential in

response to the bare Coulomb potential of an added charge of -1 electron in two positions is shown: (1) in the bond center, (2) in the interstitial site. This approximately follows the local charge density: as is intuitively

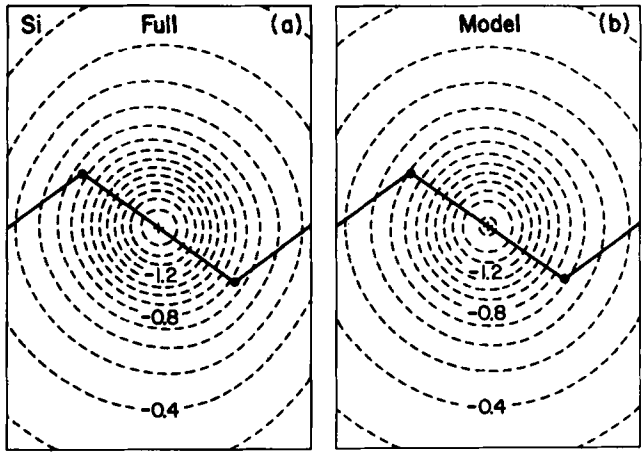


Fig. 1. Static screening potential around an electron added at the bond center as calculated (a) with the ab initio *dielectric matrices and (b) with the model Eq. (5). Plot is in the* (110) *plane with contour interval of 0.1 Ry. Bond chain indicated schematically.*

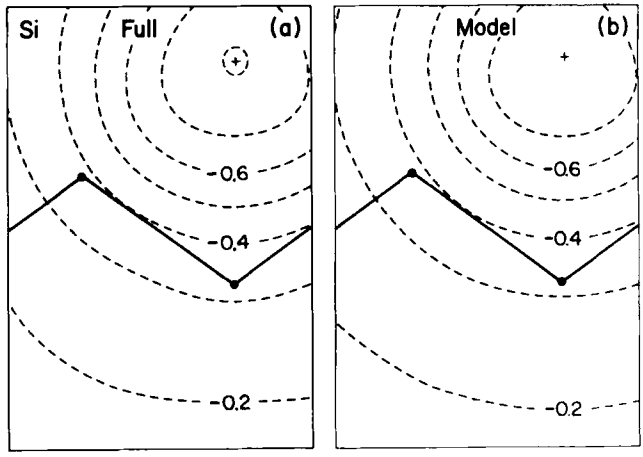


Fig. 2. Same as Fig. 1 for an electron added at the interstitial site.

clear, more electrons nearby lead to more complete screening locally. This is exploited in a model we recently introduced [18]. The model consists in taking the screening potential around an added electron at $\mathbf{r}'$ to be the same as for a homogeneous medium of the local density at $\mathbf{r}'$. For semiconductors, the model for the homogeneous medium must include the finite dielectric constant. Thus ϵ_0 is the single parameter entering the model. With the necessary symmetry,

$$V_{scr}(\mathbf{r}, \mathbf{r}') = \frac{1}{2}\left(V_{scr}^{hom}(\mathbf{r}-\mathbf{r}'; r_s(\mathbf{r}')) + V_{scr}^{hom}(\mathbf{r}'-\mathbf{r}; r_s(\mathbf{r}))\right). \tag{5}$$

Upon Fourier transformation, for a given $\mathbf{q}$, each row of ϵ^{-1} can be obtained by a fast Fourier transform. It is quite easy to implement this calculation, and the computer time required is significantly less than for the full calculation of the dielectric matrix. This model quantitatively reproduces the screening potential around an added electron in bulk materials such as Si as illustrated in Figs. 1 and 2 where part (b) shows the results of the model Eq. (5).

For the semiconductors and insulators, the random phase approximation has been used to obtain ϵ^{-1}. However, corrections to ϵ^{-1} due to exchange-correlation effects may be introduced within the LDA [16]. This was found to be negligible for the band gaps of the semiconductors but in the case of simple metals, these effects are more important. The role of exchange-correlation effects is illustrated for the case of Li in Fig. 3. In contrast to the insulators and semiconductors, the crystalline local fields give a small change while the exchange-correlation effects lead to significantly better screening at long wavelengths.

Once the screening response has been established, the self energy operator may be evaluated and Eq. (1) solved. In the Green's function Eq. (4), the quasiparticle wavefunctions and spectrum are initially approximated by the solutions of the LDA Kohn-Sham equations. Then the spectrum is updated. The LDA wavefunctions are adequate [9,10]. Equation (1) can be solved perturbatively to first order in the difference $\Sigma - V_{xc}$. (V_{xc} is the LDA exchange-correlation potential.) Second order contributions are negligible. For bulk materials, numerical precision of 0.1 eV can be typically achieved for band gaps and dispersions. In the

larger scale surface systems discussed below, the precision of the results for the empty states is slightly worse, about 0.2 eV, largely due to convergence with respect to basis set size in the underlying LDA calculation. We stress that the magnitudes of the separate matrix elements of Σ and V_{xc} are comparable and of order 10 eV for the semiconductors. However, the difference which enters the first order correction to the gaps is of order 1 eV. To obtain 0.1 eV accuracy for the gaps, the self energy operator must be calculated with a relative precision of about 1%. (Strictly speaking the difference between two matrix elements of the self energy enters the calculations here e.g. the conduction band edge and the valence band edge for the minimum gap.) This requirement imposes stringent constraints in the numerical treatment of the problem and makes it quite difficult to find simplified models for the self energy operator that are sufficiently accurate.

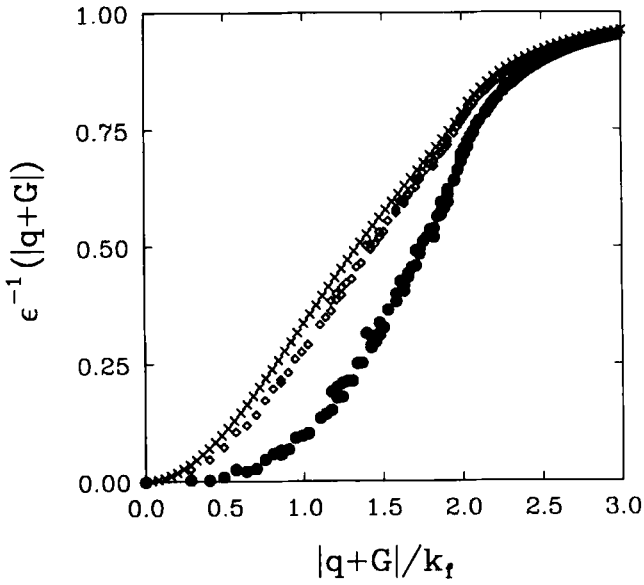


Fig. 3. Diagonal elements of ϵ^{-1} as a function of momentum for the Lindhard model (+), the full crystalline calculation in RPA (◇) and including exchange-correlation effects (●).

3. APPLICATIONS

3.1. Bulk Semiconductors and Insulators

The quasiparticle energies have been calculated using the present self energy approach for materials spanning a wide range of gap size and ionicity. The results for the minimum gap are shown in Table I in comparison to the LDA gap as well as experiment [19]. The agreement with experiment is excellent, being within about 0.1 eV or a few percent. Although the self energy approach is not formulated as a "correction" to the LDA eigenvalues, it is still interesting to isolate the physically interesting effects which lead to an improved gap. The GW self energy operator Σ captures three features not properly found in the local exchange-correlation potential. First, it is spatially a non-local operator. Thus it is sensitive to the nodal structure of the wavefunctions and there is an abrupt change in the character of the wavefunctions upon crossing the gap. This aspect is already present in the bare exchange operator (Hartree-Fock theory) which yields a gap that is considerably larger than experiment. The difference is due to the correlation contribution to Σ. In the GW approximation, this naturally breaks into two parts: (i) the exchange term from Hartree-Fock theory is dynamically screened; (ii) there is a "Coulomb hole" term arising from the interaction of the added electron with the dynamical screening cloud it induces. Second, the inclusion of local fields in the screening is crucial for the strongly

E_g	LDA	QP	Expt
diamond	3.9	5.6	5.48
Si	0.52	1.29	1.17
Ge	<0	0.75	0.744
LiCl	6.0	9.1	9.4

Table I. Comparison of the calculated minimum gap using the present self energy approach (QP) to the results from the LDA eigenvalues and experiment [19].

inhomogeneous electron distributions found in semiconductors and insulators. The strong spatial variation in the screening response is illustrated in Figs. 1 and 2. If neglected, the calculated gaps are significantly too small [9]. Local fields in the Coulomb hole term in particular open the gap as can be seen with reference to Figs. 1 and 2 and noting that the valence band edge wavefunctions are concentrated in the bond regions while the conduction band edge states of Si are primarily in the interstitial region. Thirdly, adequate account must be taken of the frequency dependence of the screening. A static approximation leads to gaps which are generally too big [9]. Retardation in the interaction of the added electron with its screening cloud leads to a reduction of the Coulomb hole term.

We turn now to a broader examination of the quasiparticle energies in comparison to experiment. The results of the present calculation are in good agreement with data from a variety of probes: photoemission, inverse photoemission and optical response. In order to make comparison to the features in optical spectra, use is made of previous assignments of specific features to vertical transitions in particular regions of the Brillouin zone. The comparison to experiment is given in Table II (Si) and Table III (Ge). The optical transition energies agree with experiment to within about 0.1 eV. The only exception is the E_0 feature in Ge deriving from the zone center s-like conduction band ($\Gamma_{2'c}$ state). However, as discussed in Ref. 9, this state is strongly influenced by the exchange with the core states, which is not treated exactly here since an ionic pseudopotential generated using the LDA is employed. The data from inverse photoemission also agree with the present calculation within the precision available. Comparison to photoemission data for the occupied states shows that the calculated bands are somewhat too narrow. However, as illustrated in Fig. 4, the overall results for the valence band dispersion agree well with experiment [20]. Finally, the model given by Eq. (5) has been used to calculate the quasiparticle energies. As can be seen from Tables II and III, the results agree with those from the full calculation within about 0.1 eV.

In Fig. 5, we make a direct comparison between the calculated

	QP	QPM	Expt.
E_g	1.29	1.16	1.17
$\Gamma_{1v} \rightarrow \Gamma_{25'v}$	12.04	12.45	12.5±.6
$\Gamma_{25'v} \rightarrow \Gamma_{15c}$	3.35	3.32	3.4
$\Gamma_{25'v} \rightarrow \Gamma_{2'c}$	4.08	4.14	4.2
$L_{3'v} \rightarrow \Gamma_{25'v}$	1.27	1.31	1.2±.2, 1.5
$\Gamma_{25'v} \rightarrow L_{1c}$	2.27	2.22	2.1, 2.4±.15
$\Gamma_{25'v} \rightarrow L_{3c}$	4.24	4.19	4.15±.1
$L_{3'v} \rightarrow L_{1c}$	3.54	3.53	3.45
$L_{3'v} \rightarrow L_{3c}$	5.51	5.50	5.50

Table II. Comparison of calculated band gaps and dispersions for bulk Si based on the full quasiparticle theory (QP) and using the model Eq. (5) for the dielectric matrices (QPM) to experimental data from Ref. 19.

	QP	QPM	Expt.
E_g	0.75	0.60	0.744
$\Gamma_{6v} \rightarrow \Gamma_{8v}$	12.86	13.06	12.6, 12.9±.2
$\Gamma_{8v} \rightarrow \Gamma_{7c}$	0.71	0.62	0.89
$\Gamma_{8v} \rightarrow \Gamma_{6c}$	3.04	2.91	3.006
$\Gamma_{8v} \rightarrow \Gamma_{8c}$	3.26	3.13	3.206
$L_{6v} \rightarrow \Gamma_{8v}$	1.61	1.64	1.4±.3
$L_{4,5v} \rightarrow \Gamma_{8v}$	1.43	1.46	
$\Gamma_{8v} \rightarrow L_{6c}$	4.33	4.19	4.3±.2, 4.2±.1
$\Gamma_{8v} \rightarrow L_{4,5c}$	4.43	4.29	
$\Gamma_{8v} \rightarrow L_{6c}$	7.61	7.43	7.8±.6, 7.8±.1

Table III. Same as Table II for Ge.

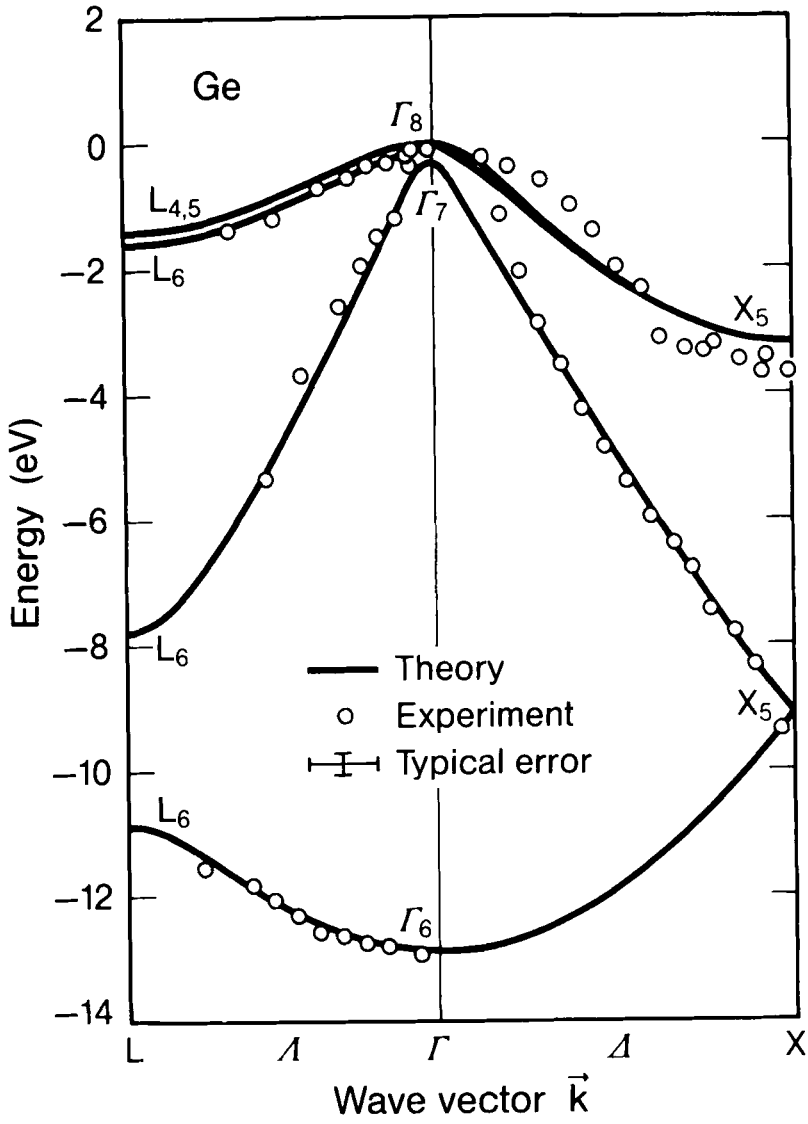


Fig. 4. Comparison of calculated quasiparticle valence band dispersions along the Λ and Δ lines in the Brillouin zone for Ge to results from angular-resolved photoemission measurements from Ref. 20.

quasiparticle energies and the LDA eigenvalues for the materials considered here. As can be seen, the difference (or "correction" to the LDA) is a relatively smooth function of energy for the valence and conduction band states separately but that there is a sharp jump at the gap. This is reflective of the discontinuity in the true exchange correlation potential [6-8]. From the point of view of "correcting" the LDA eigenvalues, the deviation is seen to be material dependent as well as energy dependent showing that a simple rigid shift of conduction bands relative to the valence band edge is inadequate. It is interesting to recall that the highest occupied eigenvalue in the true DFT should give the correct removal energy (quasiparticle energy). Therefore, the extent to which the difference is non-zero for the valence band edge state in Fig. 5 is an indirect indication of the accuracy of the LDA for that material. Interestingly, there is a clear trend of decreasing deviation with increasing metallicity (increasing dielectric constant). This supports the common notion that the LDA is better for more metallic (less inhomogeneous) systems.

3.2. Simple Metals

Recent angular-resolved photoemission experiments on Na, the prototypical simple metal, suggest a significant narrowing of the band width in comparison to free electron (or LDA) bands [5]. In particular, the bottom of the band is found to be 2.5-2.65 eV below the Fermi level as compared to the free electron value of 3.2 eV. The original GW calculations [1,15] for the electron gas of the relevant density suggested a narrowing of at most 0.3 eV. We find [11] that it is important to include the exchange-correlation effect in the screening. When this is done using a LDA dielectric function, together with self consistent update of the spectrum in the Green's function, a narrowing of 0.7 eV is calculated. Use of a Hubbard model dielectric function yields slightly less narrowing. The final position of the bottom of the band in Na is in excellent agreement with experiment. Our quasiparticle energies for Na are also consistent with X-ray absorption edge measurements which are sensitive to the position of the empty density of states features deriving from the

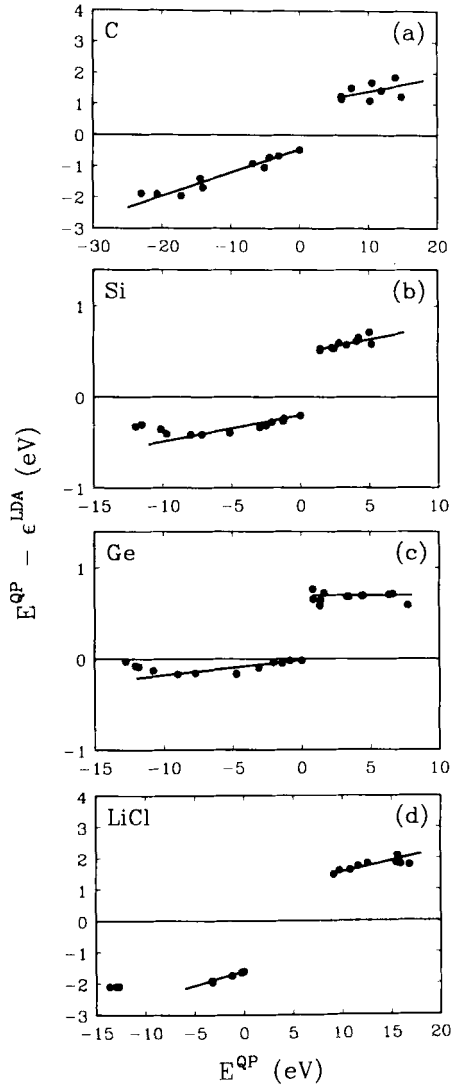


Fig. 5. Difference between the calculated quasiparticle energy (E^{qp}) and the LDA eigenvalue (ϵ^{LDA}) as a function of quasiparticle energy for several states at high symmetry points in the Brillouin zone for (a) diamond, (b) Si, (c) Ge, and (d) LiCl. Lines are a guide to the eye and the zero for E^{qp} is the valence band edge.

gap at the zone face [21]. Similar narrowing of the occupied states is calculated for other simple metals. The increased importance of exchange-correlation effects is probably due to the generally lower average electron density in these materials and the larger role played by smaller q screening. In the semiconductors and insulators, the local fields play the dominant role.

3.3. Semiconductor Interfaces and Short Period Superlattices

Advances in materials growth techniques have made possible experiments on semiconductor superlattice structures alternating just a few monolayers of each type of semiconductor. The electronic structure of these materials can be calculated directly using bulk techniques. Novel features are expected in the optical response due to the lowered symmetry and artificial periodicity along the growth direction. Furthermore, since the effect of the interface is screened out very quickly, these superlattices can be used to study the properties of the semiconductor heterojunction, in particular the band offsets. Both problems require calculation of the quasiparticle energies. Here we focus on the band offset problem and note that the self energy approach has played an important role in the interpretation of optical spectra for short period Si/Ge strained layer superlattices [13].

The self consistent LDA superlattice calculation yields the ground state charge density and hence the dipole contribution to the valence band offset across the interface. Since the LDA charge density has proven accurate in comparison to experiment for semiconductors, we expect no further significant correction to this term in the band offset. Previous calculations of this type for the offset have then used the LDA valence band edge energy in the respective bulk materials relative to the average local potential to derive the final valence band offset [22]. However, Fig. 5 clearly shows that there are material dependent self energy corrections to the valence band edge. We have calculated these self energy corrections for the experimentally well characterized case of GaAs/AlAs (001) interfaces.

The self energy operator and LDA exchange-correlation potential

are localized operators, so that the absolute value of the difference is well defined and it is the bulk value which properly enters the band offset. Thus, the valence band offset can be written

$$\Delta E_v = \Delta E_v^{LDA} + \delta_{vbm}, \tag{6a}$$

$$\delta_{vbm} = (\Sigma - V_{xc})^{vbm}_{GaAs} - (\Sigma - V_{xc})^{vbm}_{AlAs}. \tag{6b}$$

The result of the self energy calculation is $\delta_{vbm} = 0.12 \pm 0.02$ eV. When combined with the $\Delta E_{vbm}^{LDA} = 0.41$ eV, the final offset of $\Delta E_v = 0.53 \pm 0.05$ eV follows which agrees well with the present experimental range of 0.53 - 0.56 eV [23,24]. The self energy contribution to the valence band offset in this system is about 30% of the LDA value and is quantitatively important. The many-body correction can be traced to the relatively tighter p-orbitals in AlAs in comparison to GaAs which leads to a deeper exchange hole.

3.4. Semiconductor Surfaces

A monolayer of As on Si(111) and Ge(111) surfaces has been shown to form a nearly ideal 1×1 surface with As substituting for the outer half of the surface double layer [25,26]. The surface is stabilized by the fully occupied lone pair state on the As at the surface. These surfaces present an excellent prototype for the self energy approach. This application further demonstrates the usefulness of combining LDA calculations to obtain the ground state configuration with quasiparticle calculations for the spectroscopic properties.

The position of the As layer has been relaxed with respect to the bulk Si and Ge by minimizing the LDA total energy. The resulting position of the As layer agrees very well with the X-ray standing wave determination for the Si case [26]. The results of the self energy calculation for Ge(111):As are summarized in Fig. 6. The As lone pair band is occupied and resonant through most of the Brillouin zone. In comparison to the LDA, the dispersion is enhanced. The results of the calculation are compared to angular resolved photoemission data [25] in Fig. 7 showing excellent agreement. Results are similar for Si(111):As. In addition, there is an empty surface state in the gap near $\bar{\Gamma}$. The position

of the empty state was predicted using the self energy approach and subsequently confirmed quantitatively with scanning tunneling spectroscopy [27]. The calculated gap between empty and occupied surface states at $\bar{\Gamma}$ is 0.79 eV (2.22 eV) in comparison to experiment 0.8 eV (1.9 eV) for Ge(111):As (Si(111):As).

It is again interesting to examine the difference between the quasiparticle energies and the LDA eigenvalues. This is plotted for the Ge(111):As surface in Fig. 8a in analogy to Fig. 5 for the bulk case. The bulk derived states behave exactly as in Fig. 5 for bulk Ge. However, the occupied surface states show a distinct correction and the $\bar{\Gamma}$ surface states interpolate through the jump seen at the gap for the bulk states. This intriguing behavior suggests a simple physical picture. The surface states derive weight both from bulk valence and conduction bands. One can write at $\bar{\mathbf{k}}$ in the surface Brillouin zone

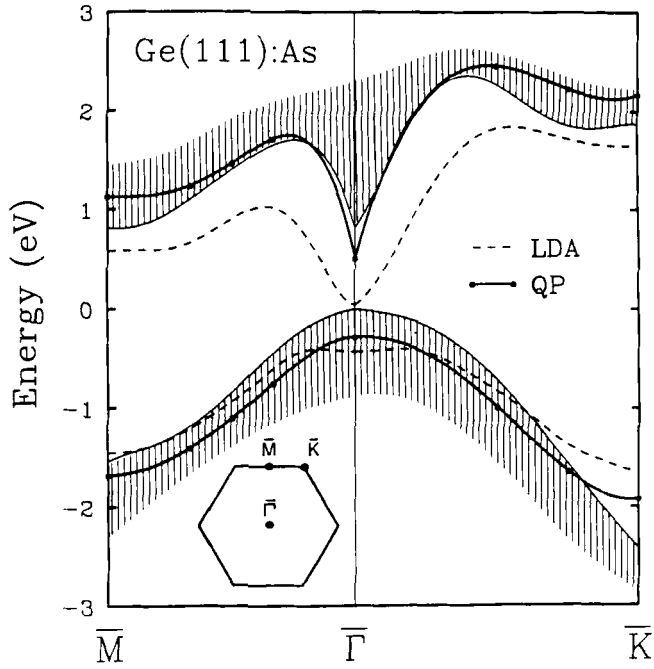


Fig. 6. Calculated surface band energies plotted against the bulk projected bands for the Ge(111):As surface.

$$<s\bar{\mathbf{k}}\,|\,\Sigma - V_{xc}\,|\,s\bar{\mathbf{k}}> = \sum_b |<s\bar{\mathbf{k}}\,|\,b\bar{\mathbf{k}}>|^2 <b\bar{\mathbf{k}}\,|\,\Sigma_{bulk} - V_{xc,\,bulk}\,|\,b\bar{\mathbf{k}}> + <s\bar{\mathbf{k}}\,|\,\Delta\Sigma - \Delta V_{xc}\,|\,s\bar{\mathbf{k}}>, \quad (7)$$

where the surface state (s) matrix elements are broken into two pieces: one from the projection of the surface state on the complete set of bulk states (b) and the remainder. In the first term, off-diagonal terms are taken to be approximately zero [9]. The projections $|<s\bar{\mathbf{k}}\,|\,b\bar{\mathbf{k}}>|^2$ are plotted in Fig. 8b in spectral form with Gaussian broadening for presentation. If the second term in Eq. (7) is neglected, the "correction" can be calculated for the $\bar{\Gamma}$ surface states within about 0.1 eV. The intermediate values seen in Fig. 8a clearly follow from the mixed conduction and valence band spectral weight in these surface states and the jump in the bulk "correction." However, the surface state at K also exhibits a mixture of valence and conduction band character, yet the difference in Fig. 8a is more negative than for a bulk state. Quantitative analysis shows that the second term in Eq. (7) must contribute about -0.3 eV in this case. This shows that an accurate description of the surface state self energy would require a model for the second term in Eq. (7).

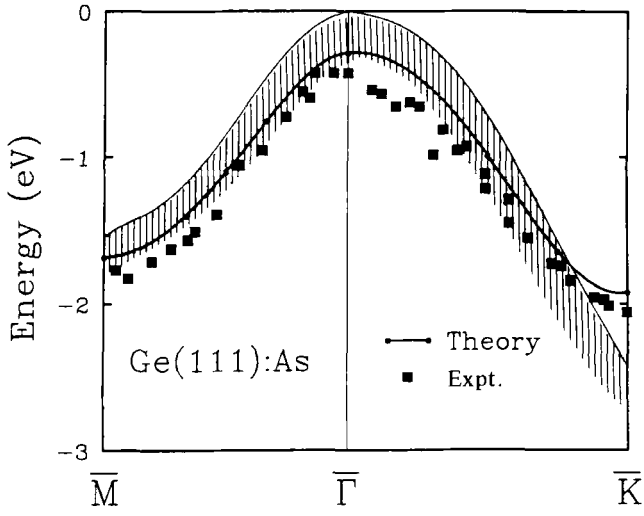


Fig. 7. Comparison of calculated occupied surface band energies to results of angular resolved photoemission experiments [25] for Ge(111):As.

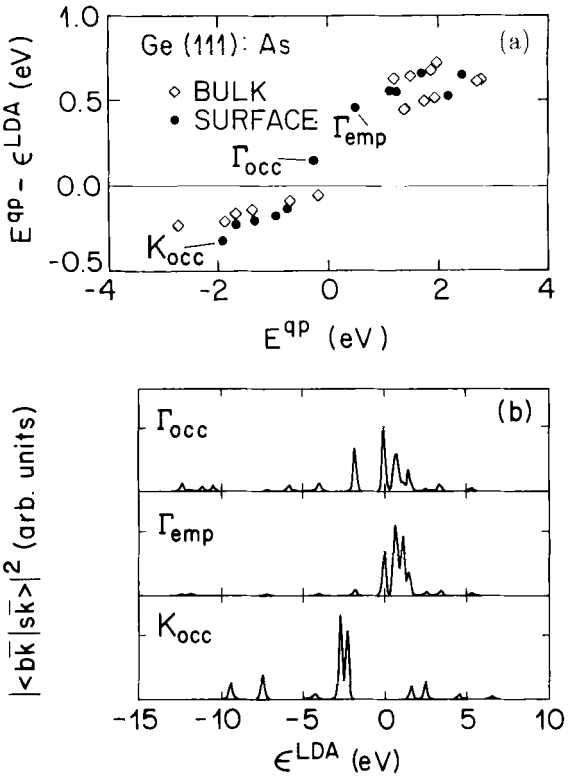


Fig. 8. (a) Difference between the calculated quasiparticle energy and the LDA eigenvalue as a function of quasiparticle energy near the gap region for the Ge(111):As surface. (b) Projected weight of selected surface states onto bulk states. Energy of bulk valence band edge taken as zero.

4. CONCLUDING REMARKS

We have briefly reviewed the successful application of a many-body theory to the calculation of quasiparticle energies in real materials. The accuracy of the present self energy approach has been demonstrated for several diverse cases showing that a predictive theory is now available for treating spectroscopic properties of materials. Open questions include the possible role of vertex corrections (higher order terms in the perturbation expansion for the self energy) as well as the feasibility of

finding a model for the self energy operator which could be applied to real materials with the relative ease associated with standard LDA band calculations. The relative accuracy with which the self energy must be obtained in order to achieve reliable quasiparticle energies is a stringent requirement for such models.

ACKNOWLEDGEMENTS

Much of the work reviewed here was done while one of us (MSH) was in the Physics Department at the University of California, Berkeley. The work there was supported by National Science Foundation Grant No. DMR8818404 and by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division of the U. S. Department of Energy, under Contract No. DE-AC03-76SF00098.

REFERENCES

[1] L. Hedin and S. Lundquist, Solid State Phys. **23**, 1 (1969).
[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] *Theory of the Inhomogeneous Electron Gas,* edited by S. Lundqvist and N.H. March (Plenum, New York, 1983) and references therein.
[4] L.J. Sham and W. Kohn, Phys. Rev. **145**, 561 (1966).
[5] E. Jensen and E.W. Plummer, Phys. Rev. Lett. **55**, 1912 (1985); I.-W. Lyo and E.W. Plummer, Phys. Rev. Lett. **60**, 1558 (1988).
[6] L.J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983); Phys. Rev. B **32**, 3883 (1985).
[7] J.P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).
[8] R.W. Godby, M. Schlüter and L.J. Sham, Phys. Rev. Lett. **56**, 2415 (1986); Phys. Rev. B **37**, 10159 (1988).
[9] M.S. Hybertsen and S.G. Louie, Phys. Rev. Lett. **55**, 1418 (1985); Phys. Rev. B **34**, 5390 (1986).

[10] R.W. Godby, M. Schlüter and L.J. Sham, Phys. Rev. B **35**, 4170 (1987).
[11] J.E. Northrup, M.S. Hybertsen and S.G. Louie, Phys. Rev. Lett. **59**, 819 (1987).
[12] S.B. Zhang, *et al.*, Solid State Comm. **66**, 585 (1988).
[13] M.S. Hybertsen and M. Schlüter, Phys. Rev. B **36**, 9683 (1987).
[14] M.S. Hybertsen and S.G. Louie, Phys. Rev. Lett. **58**, 1551 (1987); Phys. Rev. B **38**, 4033 (1988).
[15] L. Hedin, Phys. Rev. **139**, A796 (1965).
[16] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **35**, 5585 (1987); *ibid.*, 5602 (1987); references therein.
[17] W. von der Linden and P. Horsch, Phys. Rev. B **37**, 8351 (1988).
[18] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **37**, 2733 (1988).
[19] *Zahlenwerte und Funktionen aus Naturwissenshaften und Technik*, in Vol. III of Landolt-Bornstein (Springer, New York, 1982), pt. 17a; see also Ref. 9.
[20] A.L. Wachs, *et al.*, Phys. Rev. B **32**, 2326 (1985).
[21] P.H. Citrin, *et al.*, Phys. Rev. Lett. **61**, 1021 (1988).
[22] C.G. Van de Walle and R.M. Martin, J. Vac. Sci. Technol. B **4**, 1055 (1986).
[23] P. Dawson, K.J. Moore and C.T. Foxon, *Quantum Well and Superlattice Physics,* (Bay Point), Proc. of SPIE, **792**, ed. G.H. Döhler and J.N. Schulman (SPIE, Washington, 1987), p.208.
[24] D.J. Wolford, *Proc. of the 18th Int. Conf. on the Physics of Semiconductors,* Stockholm (1986), p. 1115; private communication.
[25] R. D. Bringans, *et al.*, Phys. Rev. Lett. **55**, 533 (1985); R. I. G. Uhrberg, *et al.*, Phys. Rev. B **35**, 3945 (1987).
[26] J.R. Patel, *et al.*, Phys. Rev. **36**, 7715 (1987).
[27] R.S. Becker, *et al.*, Phys. Rev. Lett. **60**, 116 (1988); unpublished.

RESPONSE FUNCTIONS AND NONLOCAL APPROXIMATIONS

David C. Langreth

Department of Physics and Astronomy
Rutgers University, Piscataway, New Jersey 08855-0849

S. H. Vosko

Department of Physics
University of Toronto, Toronto, Ontario M5S 1A7

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

1. INTRODUCTION

Since the pioneering work of Hohenberg and Kohn /1/ and Kohn and Sham /2/, two principal methods have been used for constructing nonlocal density functionals for the exchange and correlation energy. The first is to consider the density-density correlation function or exchange-correlation hole of the nonuniform system, and to approximate it via various ad hoc assumptions, while making sure that various exact sum rules and symmetries remain satisfied. The second is to make a priori calculations of the linear response functions of the uniform system and then to extrapolate the results to encompass the degree of nonuniformity actually encountered in the physical systems of interest. It was this latter approach that was developed in the Hohenberg and Kohn paper. In particular it was emphasized that the linear response determined the nonlocal corrections in two limiting cases: (i) Systems of slowly varying density for which only the long wavelength behavior of the density response is required (gradient expansions); and (ii) Systems of almost constant density with small oscillations of arbitrary wavelength which for implementation require the density response for all wavelengths. Real systems do not satisfy either of these limits and moreover because of the unexpected rapid variation with wavevector of the pertinent parts of the density response (as suggested by the work of Langreth and Perdew /3/, Langreth and Mehl /4/, and recently directly proved by the authors /5/), it appears that practical methods must in some sense bridge these two regimes.

1.1. The Situation prior to 1980

This article focuses on the connection between response functions and nonlocal approximations, with emphasis on the gradient-type expansion methods and some new extensions. Such expansions had fallen into disrepute prior to 1980 for a variety of reasons. The ability of the gradient expansion to correct the local density approximation (LDA) has usually been assessed by applying it to atomic systems for exchange and correlation separately /6/. It was recognized early that the LDA for correlation was too negative by a

factor of about 2. In a classic work, Ma and Brueckner /7/ investigated the effect of the gradient correction by evaluating it in the high density limit and discovered two unexpected results: (i) its dependence on e^2 is linear, the same as for exchange and (ii) the gradient term overestimates the correction needed to reproduce the semi-empirical correlation energy by a factor of five /8/. Thus a naive application of their result would produce positive correlation energies—an impossible result. Ma and Brueckner did not give a microscopic explanation of the source of this anomaly, but did note that there were two length scales involved in the correlation problem, and used this argument to combine their theoretical results with the semi-empirical values for the correlation energies of a number of atoms to produce a formula, which could be called a "generalized gradient expansion" for the correlation energy. This formula has not been found to be useful. Soon afterwards Herman et al. /9/ investigated the gradient correction for exchange where, in distinction to correlation, the LDA is not sufficiently negative. They determined a semi-empirical coefficient for the gradient correction by fitting to a number of atoms. This form when applied to a number of systems was found to improve densities (e. g. x-ray form factors). Sham/10/ then made a first principles evaluation of the exchange coefficient and found it to be $\sim \frac{1}{3}$ of the semi-empirical value /11/.

Thus the situation pre–1980 may be summarized as follows: For exchange and correlation treated separately the gradient corrections have the proper signs to improve their respective LDA values. Quite reasonable results were obtained for the exchange part; however for correlation the corrections were anomalously large and put the entire procedure in doubt.

1.2. Recent Progress

The real power of density functional theory is its capacity to treat exchange and correlation on the same footing, in contrast to conventional quantum mechanical methods where the inclusion of correlation requires elaborate procedures. Thus, in some respects the more pertinent question to pose is what happens when exchange and correlation are treated together? It is found that, like exchange alone, the LDA does not give sufficiently negative values (i. e. the errors in the individual parts tend to cancel, but not completely) and unfortunately the gradient correction is now positive because of the spuriously large correlation correction discussed above /12/. The important progress that has been made in recent years is to provide a microscopic understanding of the source of this spurious result. In particular it has been shown that:

- The nonlocal part of the density response of the uniform system (an interacting electron gas) to a potential of wavevector q, rapidly deviates from its power series expansion to order q^2 even when q is very small /5/. This confirms by direct calculation the claims that a number of authors including one of us /3, 4, 13, 14/ have made, based on more indirect methods. It also shows that the simple gradient approximation /7, 15, 16/, which is based on the validity of this second order power series expansion over a reasonable range of q [up to $\sim k_F$], is not useful for the density variations encountered in real systems.

In addition, the above concepts have been very recently tested on extended systems:

1. There is now verification, through the recent calculations of Bagno, Jepsen, and Gunnarsson /17/, that the gradient approximation is worse than the local density approximation for the ground state properties (cohesive energy, equilibrium density, and bulk modulus) of a range of bulk metals.

2. Electronic structure calculations /17, 18, 19/ have also been made with approximate functionals /4, 20, 21, 22, 23/ that were explicitly designed to take into account the rapid variation mentioned in the bulleted item above. These calculations systematically give a significant improvement over the local density approximation for the same equilibrium properties of a wide range of bulk metals.

It is not the purpose of this article to promote the use of any particular functional. For example, no functional we know of makes a substantial reduction in the well-known error of the local density approximation for calculating inter-configurational energies/21, 24, 25/ or the correlation energy contribution to the electron affinities in which a 3d electron is removed in the iron series /24/. It is our purpose, however, to argue that the rapid variation mentioned above is a property that all reasonable functionals should reflect, and to suggest that this property, which is incorporated in the approximate functionals mentioned above /4, 20, 21, 22, 23/, be incorporated in the development of still better functionals in the future. We will therefore focus on the behavior of the density response function, and how it is related to the development of nonlocal exchange and correlation energy functionals.

2. DENSITY FUNCTIONAL THEORY AND LINEAR RESPONSE

As mentioned in the introduction, a large number of the density functionals that have been proposed and used in practice, have their origins in extrapolations of results obtained in linear response starting with a uniform interacting electron gas. In this section, we will use elementary, but rigorous, arguments to obtain some of the relations that have been found useful over the years.

2.1. Density Response Functions

As we will consider only single component systems, the terms "charge density" and "particle density" will be used interchangeably, and correspondingly the fundamental particle (an electron) has unit charge. Consider an external charge with the same sign as the electron

$$n^{\mathrm{ext}}(\mathbf{r}) = \sum_{\mathbf{q}} n_{\mathbf{q}}^{\mathrm{ext}} e^{i\mathbf{q}\cdot\mathbf{r}}, \tag{1}$$

causing an external potential, whose Fourier transform is $V_{\mathbf{q}}^{\mathrm{ext}} = (4\pi/q^2)n_{\mathbf{q}}^{\mathrm{ext}}$, acting on a uniform system whose electronic density is n_0. If we let $n(\mathbf{r})$ be the electronic density of the system and treat V^{ext} as a small perturbation, then

$$n(\mathbf{r}) = n_0 + n^{\mathrm{ind}}(\mathbf{r}), \tag{2}$$

and

$$n^{\mathrm{ind}}(\mathbf{r}) = \sum_{\mathbf{q}} n_{\mathbf{q}}^{\mathrm{ind}} e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{3}$$

Associated with $n_{\mathbf{q}}^{\mathrm{ind}}$ is the potential $V_{\mathbf{q}}^{\mathrm{ind}} = (4\pi/q^2)n_{\mathbf{q}}^{\mathrm{ind}}$. The dielectric response function $\epsilon(q)$ can be defined in terms of the ratio of this induced potential and the external potential:

$$V_{\mathbf{q}}^{\mathrm{ind}} = \left[\frac{1}{\epsilon(q)} - 1\right] V_{\mathbf{q}}^{\mathrm{ext}}. \tag{4}$$

In a similar manner, one defines the density response function or susceptibility as the density component induced by the corresponding component of the *external* potential:

$$n_{\mathbf{q}}^{\mathrm{ind}} = \chi(q) V_{\mathbf{q}}^{\mathrm{ext}}, \tag{5}$$

so that according to (4)

$$\frac{1}{\epsilon(q)} - 1 = \frac{4\pi}{q^2}\chi(q). \tag{6}$$

Here, we shall deal most directly with the density response to the total electric potential, which we denote /26/ by $\chi_{sc}(q)$, and which is simply $\chi(q)\epsilon(q)$, so that

$$\epsilon(q) = 1 - \frac{4\pi}{q^2}\chi_{sc}(q). \tag{7}$$

In a diagrammatic formulation (see Fig. 1), the quantity $\chi_{sc}(q)$ is the irreducible polarization part, which is often denoted by $\Pi(q)$ or $-\Pi(q)$.

2.2. Density Functional Theory and Energy Response Functions

In this section we review the notions introduced by Hohenberg and Kohn /1/ with respect to density functional theory in the linear response regime. The total energy functional E[n] takes the form

$$E[n] = F[n] + \int d^3r\, n(\mathbf{r}) V^{ext}(\mathbf{r}). \tag{8}$$

where $F[n]$ is a universal functional of n. If we suppose as in (2) that

$$n(\mathbf{r}) = n_0 + \delta n(\mathbf{r}), \tag{9}$$

where δn is the density induced by some local external potential, not necessarily V^{ext}. Using (9) and expanding to second order in $\delta n(\mathbf{r})$ gives (8)

$$F[n] = F(n_0) + \int d^3r_1 \int d^3r_2\, K_t(\mathbf{r_1} - \mathbf{r_2})\delta n(\mathbf{r_1})\delta n(\mathbf{r_2}), \tag{10}$$

thus defining K_t, whose only hidden functional dependence is in n_0. There is no term linear in $\delta n(\mathbf{r})$ because of the translational invariance of the uniform system plus number conservation $[\int d^3r\, \delta n(\mathbf{r}) = 0]$. Upon Fourier transformation, (10) then becomes

$$E[n] = E[n_0] + \sum_{\mathbf{q}} K_t(q)\delta n_{\mathbf{q}}\delta n_{-\mathbf{q}} + \sum_{\mathbf{q}} V^{ext}_{\mathbf{q}}\delta n_{-\mathbf{q}}. \tag{11}$$

We call K_t the total energy response function [see Eq. (14) below]. It is related to the response function K, defined by Hohenberg and Kohn /28/ through $K_t(q) = K(q) + 2\pi/q^2$. As first shown by Hohenberg and Kohn, such response functions have a simple connection to the density response function.

2.3. Connection between Energy Response and Density Response

The energy response function K_t and the density response function χ are rather simply related. We will derive this connection in two different ways: first from the Hohenberg-Kohn theorem, and second using a simple, but

rigorous, classical argument. The original derivation by Hohenberg and Kohn was somewhat different.

According to the Hohenberg-Kohn theorem, the quantity $n_{\mathbf{q}}^{\mathbf{ind}}$, that is the value of $\delta n_{\mathbf{q}}$ actually induced by the potential $V_{\mathbf{q}}^{\mathbf{ext}}$ is given by the value of density that minimizes (11). Taking the variation gives

$$-2K_t(q)n_{\mathbf{q}}^{\mathbf{ind}} = V_{\mathbf{q}}^{\mathbf{ext}}. \tag{12}$$

A comparison of (12) and (5) gives the desired relation between K_t and χ:

$$K_t(q) = -\frac{1}{2\chi(q)}. \tag{13}$$

Finally, substitution back in (11) yields

$$E[n] = E[n_0] - \sum_{\mathbf{q}} K_t(q) n_{\mathbf{q}}^{\mathbf{ind}} n_{-\mathbf{q}}^{\mathbf{ind}}. \tag{14}$$

We now derive these key equations (12–13) via classical electrostatics. Actually it is more convenient to take (14) as the definition of K_t and derive (11) instead. The work W required to build up an external charge is

$$W = \frac{1}{2}\sum_{\mathbf{q}} V_{-\mathbf{q}}^{\mathbf{ind}} n_{\mathbf{q}}^{\mathbf{ext}} = \frac{1}{2}\sum_{\mathbf{q}} n_{-\mathbf{q}}^{\mathbf{ind}} V_{\mathbf{q}}^{\mathbf{ext}} \tag{15}$$

$$= \frac{1}{2}\sum_{\mathbf{q}} \frac{1}{\chi(q)} n_{\mathbf{q}}^{\mathbf{ind}} n_{-\mathbf{q}}^{\mathbf{ind}}, \tag{16}$$

where in obtaining (16) we made use of (5). Since this work is done adiabatically, by definition of our zero frequency response functions, one must have that $E = E(n_0) + W$, so that (13) follows upon comparing (16) with (14); (12) and the equilibrium form of (11) follow immediately.

2.4. The Exchange-Correlation Functional

The most viable approximation scheme that has evolved over the years is to separate $F[n]$ into a part that corresponds to the kinetic energy of an equivalent non-interacting system, $T_s[n]$, whose explicit form is unknown but that is treated exactly by conventional quantum mechanical methods; a part that corresponds to the classical electrostatic energy; and then to do the best that one can on the rest of the functional. The poorly known part (i. e., the rest) of the functional is the so called exchange-correlation part (even though it contains some of the kinetic energy of the interacting system), and it is what we concentrate on here.

We follow Hohenberg, Kohn, and Sham /1, 2/ and define the exchange-correlation functional E_{xc} by

$$F[n] = T_s[n] + \frac{1}{2}\int d^3r_1 \int d^3r_2 \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} + E_{xc}[n], \tag{17}$$

where $T_s[n]$ is the kinetic-energy functional mentioned above. Analogously to (10) we may define an exchange-correlation kernel such that to lowest non-trivial order in δn we have

$$E_{xc}[n] = E_{xc}(n_0) + \int d^3r_1 \int d^3r_2 \, K_{xc}(\mathbf{r}_1 - \mathbf{r}_2)\delta n(\mathbf{r}_1)\delta n(\mathbf{r}_2). \tag{18}$$

If we let $\chi_0(q)$ be the density response function of a non-interacting uniform system of density n_0, then one can see from (10) and (13) that K_{xc} is given by

$$K_t(q) = -\frac{1}{2\chi_0(q)} + \frac{2\pi}{q^2} + K_{xc}(q). \tag{19}$$

The first term in (19) is the value of K_t associated with T_s in (17) while the second term in (19) corresponds to the electrostatic energy term in (17). Using (13), (6), and (7) to eliminate K_t and then χ from (19) in favor of χ_{sc}, one obtains

$$K_{xc}(q) = -\frac{1}{2}\left[\frac{1}{\chi_{sc}(q)} - \frac{1}{\chi_0(q)}\right]. \tag{20}$$

Thus, like K_t, K_{xc} is expressible in terms of density response functions.

The Kohn-Sham potential v^{KS} is that external potential which when applied to a noninteracting system yields the same density as v^{ext} does when applied to the real system. We do not discuss the very interesting questions related to the existence /29, 30, 31/ of such a potential here. In linear response, it is trivially given by

$$\chi_0(q)v_{\mathbf{q}}^{KS} = \chi(q)v_{\mathbf{q}}^{ext}. \tag{21}$$

If one separates out the exchange-correlation component of this potential via

$$v_{\mathbf{q}}^{KS} = v_{\mathbf{q}}^{ext}/\epsilon(q) + v_{\mathbf{q}}^{xc}, \tag{22}$$

then using $\chi = \chi_{sc}/\epsilon$ one obtains

$$v_{\mathbf{q}}^{xc} = \left[\frac{\chi_{sc}(q) - \chi_0(q)}{\chi_0(q)}\right]\frac{v_{\mathbf{q}}^{ext}}{\epsilon(q)}, \tag{23}$$

a result that may also be easily verified by applying the Kohn-Sham after-

thought /32/

$$v^{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}(\mathbf{r})}{\delta n(\mathbf{r})} \tag{24}$$

to Eq. (18), and using (20) along with the relation $\chi_{\mathrm{xc}} = \epsilon\chi$.

3. THE LOCAL DENSITY APPROXIMATION AND BEYOND

In the previous section, we have presented a reasonably complete description of density functional theory in linear response. We emphasize, however, that in realistic applications it is almost never possible to use such response function methods for the whole functional. What is normally done is to treat the "kinetic energy" and classical electrostatic energy parts of the functional exactly, and treat E_{xc} in LDA plus corrections. These corrections, are implicit in whatever nonlocal functional is being used, and are often based on some extrapolation or form of linear response theory, with the uniform electron gas as a starting point. Thus it is extremely important in the development of nonlocal functionals to have a good understanding of the properties of the response functions related to the exchange-correlation functional.

3.1. The Local Density Approximation

The LDA assumes that E_{xc} is locally the same in the non-uniform system as it would have been in a uniform electron gas at the same density, that is /33/

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[n] = \int d^3r\, E_{\mathrm{xc}}(n(\mathbf{r})). \tag{25}$$

Let us now extract the local and nonlocal parts from the response function K_{xc}. The nonlocal part is relevant in the construction of nonlocal functionals, even if the local part can be more accurately treated through (25). Referring to Eq. (18), we write

$$K_{\mathrm{xc}}(\mathbf{r_1} - \mathbf{r_2}) = K_{\mathrm{xc}}(q=0)\,\delta(\mathbf{r_1} - \mathbf{r_2}) + K_{\mathrm{xc}}^{\mathrm{nonloc}}(\mathbf{r_1} - \mathbf{r_2}). \tag{26}$$

Substituting (9) in (18) and expansion in powers of δn shows that the term

$$K_{\mathrm{xc}}(q=0) = \frac{1}{2}\frac{\partial^2 E_{\mathrm{xc}}(n_0)}{\partial n_0^2} \tag{27}$$

indeed is the correct coefficient for the LDA. Furthermore, since for a non-interacting system one has

$$-\frac{1}{2\chi_0(0)} = \frac{1}{2}\frac{\partial^2 T_s(n_0)}{\partial n_0^2}, \tag{28}$$

one sees from (20) that

$$-\frac{1}{\chi_{sc}(0)} = \frac{\partial^2 E(n_0)}{\partial n_0^2}, \tag{29}$$

in agreement with the usual compressibility sum rule /27/. The most important point however is that the nonlocal part in linear response can be extracted:

$$E_{xc}^{nonloc}[n] = \sum_{\mathbf{q}}[K_{xc}(q) - K_{xc}(q=0)]n_{\mathbf{q}}^{ind}n_{-\mathbf{q}}^{ind}. \tag{30}$$

This corresponds to the second case mentioned in the introduction, i. e., small oscillations of arbitrary wavelength. It is also the key equation for understanding the validity of the gradient approximation.

3.2. The Gradient Approximation

The q dependence of $K_{xc}(q)$ determines the coefficient of the lowest order term in the gradient expansion. For not too rapidly varying densities $K_{xc}(q)$ can be expanded and the lowest nonvanishing term in (30) gives

$$E_{xc}^{nonloc}[n] = \frac{1}{2}\int d^3r\, K_{xc}''(q=0)[\nabla n(\mathbf{r})]^2. \tag{31}$$

This is not quite the gradient expansion, because of the hidden dependence of $K_{xc}(q)$ on n_0. However, if the density gradients are truly small, then (31) can be written for similar systems stacked like blocks with n_0 varying slightly from system to system. The result is an equation like (31) for the composite with n_0 replaced by $n(\mathbf{r})$. Any corrections will have to be higher order in gradients. Thus one rigorously bootstraps the result (30) to a regime where the density variations can be large, provided that they are sufficiently slowly varying.

The sensitivity of the functional to the q dependence of $K_{xc}(q)$ in both of these nonlocal regimes, establishes the importance of incorporating its behavior in the development of nonlocal functionals.

4. THE q DEPENDENCE OF $K_{xc}(q)$

Our knowledge of the q dependence comes almost entirely from high density perturbation expansions. These are non-trivial because the coupling constant is not small and terms must be summed to all orders in perturbation theory. Until recently, very little was known, even in this limit. For example, it was known that the random-phase approximation (RPA) or Lindhart dielectric function becomes exact at high densities. However, the χ_{xc} in (7) is equal to χ_0 in the RPA, so that according to (20) K_{xc} vanishes in this limit. This is a consequence of the fact that at truly high densities, the kinetic energy and average Coulomb energy dominate everything. To get results of any real meaning one must go beyond RPA. Of course at $q = 0$ one can get K_{xc} from (27), because the uniform electron gas energy is quite well known /34, 35, 36/. But this is just the LDA; in this case one would use Eq. (25), rather than the lowest order approximation to it. Thus we must focus on the behavior of the $q \neq 0$ part (30).

4.1. Diagrammatic Expansion

To remove from the kernel of (30) the q dependence that determines the lowest order gradient expansion, it is useful to define the dimensionless quantity $Z(q)$ through /37/

$$K_{xc}(q) - K_{xc}(q=0) \equiv \frac{\pi}{16k_F^4} Z(q) q^2, \tag{32}$$

where $k_F^3 = 3\pi^2 n_0$. Clearly, the q dependence of $Z(q)$ provides a measure of how well the gradient expansion bridges the regimes of very slow density variations and those occurring in real physical systems. We introduce the dimensionless coupling constant λ (which is small at high density)

$$\lambda \equiv (k_{FT}/2k_F)^2 = (\pi a_0 k_F)^{-1} = \alpha r_s/\pi, \tag{33}$$

where $k_{FT}^2 = -4\pi\chi_0(0)$, $a_0 = 1/m$ (in our units) is the Bohr radius, and r_s is the usual electron-gas parameter, which is related to k_F through $\alpha r_s a_0 k_F = 1$, with $\alpha^3 = 4/9\pi$. We also note the explicit small q form of χ_0:

$$\chi_0(q) = -\frac{mk_F}{\pi^2}\left[1 - \frac{1}{3}\left(\frac{q}{2k_F}\right)^2 + \cdots\right]. \tag{34}$$

Eq. (32) can then be written

$$Z(q) = -\frac{4}{\lambda}\left(\frac{2k_F}{q}\right)^2 \chi_0(0)\,[K_{xc}(q) - K_{xc}(0)]\,. \tag{35}$$

The above provides a dimensionless way to express our final results identical to that used in Ref. /5/ and /14/.

The standard diagrammatic expansion determines χ_{sc} rather than the right side of (35) directly. We therefore also define a dimensionless quantity $\tilde{Z}(q)$, which is directly related to the χ_{sc}:

$$\tilde{Z}(q) = -\frac{2}{\lambda}\left(\frac{2k_F}{q}\right)^2 \left[\frac{\chi_{sc}(q) - \chi_{sc}(0)}{\chi_0(0)}\right]. \tag{36}$$

The constants in the above definition have been chosen so that $\tilde{Z}(q \to 0)$ is equal to the quantity called Z by Geldart and Rasolt /15, 16/. We write

$$\chi_{sc} = \chi_0 + \delta\chi, \tag{37}$$

thus defining $\delta\chi$ and

$$\tilde{Z}(q) = \tilde{Z}_0(q) + \delta\tilde{Z}(q), \tag{38}$$

where

$$\tilde{Z}_0(q) = -\frac{2}{\lambda}\left(\frac{2k_F}{q}\right)^2 \left[\frac{\chi_0(q) - \chi_0(0)}{\chi_0(0)}\right], \tag{39}$$

and

$$\delta\tilde{Z}(q) = -\frac{2}{\lambda}\left(\frac{2k_F}{q}\right)^2 \left[\frac{\delta\chi(q) - \delta\chi(0)}{\chi_0(0)}\right], \tag{40}$$

so that from (34) $\tilde{Z}_0(0) = 2/3\lambda$. The diagrams in Fig. 1 each make a contribution to $\delta\tilde{Z}(q)$ and hence to $\tilde{Z}(q)$. Although $\tilde{Z}_0(0)$ is of order λ^{-1}, the diagrams contributing to $\delta\tilde{Z}(q)$ are of order λ^0 and higher, since the diagrammatic expansion is in powers of λ.

The quantities $Z(q)$ and $\tilde{Z}(q)$ are related through (20) coupled with their definitions (35) and (36). This relationship is particularly simple in the high density limit, because then one has $\delta\chi \ll \chi_0$. Therefore substituting (20) into (35) we have

$$Z(q) \approx -\frac{2}{\lambda}\left(\frac{2k_F}{q}\right)^2 \left[\frac{\chi_0(0)\,\delta\chi(q)}{\chi_0^2(q)} - \frac{\delta\chi(0)}{\chi_0(0)}\right]. \tag{41}$$

In cases where $q \ll 2k_F$ (but not necessarily $q \ll k_{FT}$), Eq. (41) can be simplified by expanding $1/\chi_0(q)$, in powers of $q/2k_F$, using (34) and (39).

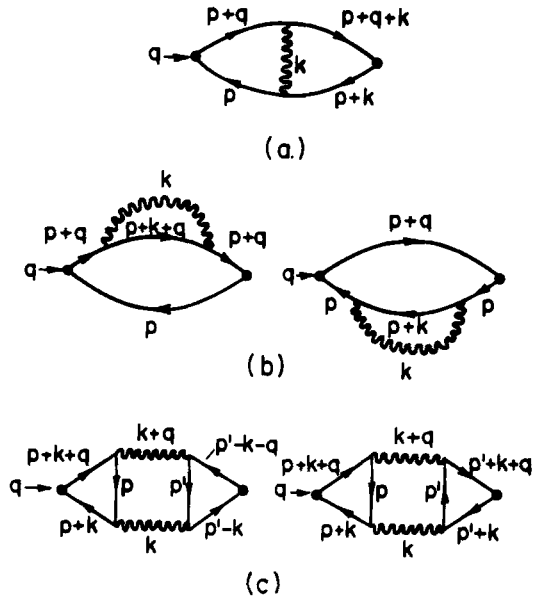


Fig. 1. Diagrams for $\delta\chi$ at high density. The wiggly lines represent random-phase approximation screened Coulomb interactions.

Upon neglecting the non-zero powers of $q/2k_F$ where appropriate, Eq. (41) then simplifies to:

$$Z(q) \approx \delta\tilde{Z}(q) - [2\delta\chi(0)/\chi_0(0)]\tilde{Z}_0(0). \tag{42}$$

The leading diagrams that contribute to $\delta\chi$ (and hence $\delta\tilde{Z}$) at high density have been known for a long time /7, 38/, and are shown in Fig. 1. Because of the linearity of (42) in $\delta\chi$ in this limit, a diagram for $\delta\chi$ (as well as $\tilde{Z}$) may equally well be regarded as a diagram for Z. In fact $\delta\chi(0)/\chi_0(0)$ is related to the compressibility of the uniform electron liquid, and has been long since well understood. In particular we know that [e. g., see Ref. /38/]

$$\delta\chi(0)/\chi_0(0) = \lambda + (1 - \ln 2)\lambda^2 + \cdots. \tag{43}$$

The first term above (proportional to λ) is the leading and only contributing term to (42) at high density. It comes from the diagrams of Fig. 1 a and b with the wiggly line replaced by a bare Coloumb interaction (exchange-only); the correlation contributions are therefore all of order λ^2 or higher and hence

negligible in this limit. Therefore, we have using (39)

$$Z(q) = \delta\tilde{Z}(q) - \tfrac{4}{3}. \tag{44}$$

In fact, if we note that as discussed above, the "4/3" term in (44) comes entirely from "exchange-only", we may write the "exchange-only" and "correlation" contributions separately as

$$Z_{\text{exchange}}(q) = \tilde{Z}_{\text{exchange}}(q) - \tfrac{4}{3}, \tag{45}$$

and

$$Z_{\text{correlation}}(q) = \tilde{Z}_{\text{correlation}}(q). \tag{46}$$

It is far from trivial to evaluate $\delta\tilde{Z}(q)$ and hence, and until recently, only a few limiting cases were known. For example as $q \to 0$ the calculations of Ma and Brueckner /7/ ($\tilde{Z}_{\text{MB}} = 1.9756$) can be combined with those of Sham /10/ ($\tilde{Z}_{\text{Sham}} = \frac{5}{9}$ so that $Z_{\text{Sham}} = -\frac{7}{9}$) to give /39/

$$\delta\tilde{Z}(0) = 2.5312, \tag{47}$$

and

$$Z(0) = 1.1978. \tag{48}$$

This is the zero q value of Z and the one that determines the coefficient of the gradient approximation through (32) and (31). It has been verified independently in many ways. For example, Rasolt and Geldart /15, 16/ have done a calculation similar to the original Ma-Brueckner one. They have also made a *finite* density calculation which approached approximately the result (48) as the density became large. Langreth and Perdew /3, 40/ using a completely different diagrammatic technique, evaluated $Z(0)$ at finite density; as the density got large, their result approached (48) to many significant figures. Finally, Langreth and Vosko /5/ have obtained the result (48) by calculating $Z(q)$ at finite q and then letting q approach zero. The message here is that the result (48) is surely correct, despite the rather large effort necessary to calculate it.

4.2. Why Is $Z(0)$ Large and Positive?

At high densities, the diagrams in Figure 1 a, b, and c are individually finite. Those in a and b are screened exchange diagrams, which make a contribution /15, 16, 41/

$$\tilde{Z}_{\text{sx}}(0) = 0.4842 \tag{49}$$

so that

$$Z_{ex}(0) = -0.8492, \tag{50}$$

while c is a fluctuation diagram, which makes a contribution /15, 16/

$$Z_{fluc}(0) = \tilde{Z}_{fluc}(0) = 2.0470. \tag{51}$$

The fluctuation involved here is a plasmon. In fact the diagrams in c have a large contribution from the zero point energy of a plasmon being scattered by the potential that produced the density variation. This interpretation was pointed out by Langreth and Perdew /3/ and verified in a detailed model by Langreth /14/. It is a large effect because the plasmon, being a long wavelength excitation, is strongly scattered by a density component of similar long wavelength.

4.3. Why Does $Z_{fluc}(q)$ Dive toward Zero as q Increases?

Physically the answer to this question is simple: plasmons exist only for wavevectors smaller or of the order of ω_p/v_F, where ω_p is the plasma frequency and v_F is the Fermi velocity. Thus as q becomes as large as the plasmon cutoff, that is

$$q \sim \omega_p/v_F = k_{FT}/\sqrt{3}, \tag{52}$$

the value of $Z_{fluc}(q)$ should become much smaller, because the density component's wavevector is no longer comparable with the wave vectors of existing plasmons. Note that in the high density limit $k_{FT}/\sqrt{3}$ is much smaller than the other wave vector where structure is expected, that is the Fermi-surface caliper $2k_F$. Even at the densities of metals such as Al, the cutoff occurs at a q, which is several times smaller than $2k_F$.

A more mathematical argument for why this must occur can be seen from the fact that the fluctuation diagram c (the dominant term at small q) is of definite sign. This is plausible simply by inspection, since both the wiggly lines and the triangular vertex functions each appear twice in each diagram. It may easily be shown rigorously simply by writing down the formal expressions for the diagram and making changes in the dummy summation variables. We combine this definiteness and the fact that the $q = 0$ value of the diagram $\delta\chi_{fluc}(0)$ is small in magnitude, with the fact that its rate of change with q^2 [which is proportional to $\tilde{Z}_{fluc}(0)$] is large in magnitude to conclude that $\tilde{Z}_{fluc}(q)$ must fall off rapidly. Specifically, we rewrite (40) for diagram c alone as

$$0 < \frac{\delta\chi_{fluc}(q)}{\chi_0(q)} \approx \frac{\delta\chi_{fluc}(0)}{\chi_0(0)} - \frac{\lambda}{2}\left(\frac{q}{2k_F}\right)^2 \tilde{Z}_{fluc}(q). \tag{53}$$

The inequality follows because $\delta\chi_{\text{fluc}}(q)$ and $\chi_0(0)$ are both negative definite. Furthermore one has /38/ that

$$\frac{\delta\chi_{\text{fluc}}(0)}{\chi_0(0)} = \frac{\lambda^2}{2}\ln\left(\frac{1}{\lambda}\right). \tag{54}$$

Therefore, (53) becomes

$$\tilde{Z}_{\text{fluc}}(q) < \left(\frac{k_{\text{FT}}}{q}\right)^2 \ln\left(\frac{2k_{\text{F}}}{k_{\text{FT}}}\right)^2. \tag{55}$$

This proves that once q becomes of the order of $k_{\text{FT}}\sqrt{\ln(2k_{\text{F}}/k_{\text{FT}})^2}$, the quantity $\tilde{Z}_{\text{fluc}}(q)$ must have begun its dive from its $q = 0$ value of 2.05 toward zero. We have for simplicity given this argument for $\tilde{Z}_{\text{fluc}}$. However, if we had used (41) instead of (40) we would have obtained a stronger inequality, like (55), but with $\tilde{Z}_{\text{fluc}}(q)$ replaced by $Z_{\text{fluc}}(q)$. Therefore if we make the reasonable assumption that the contribution from the diagrams a and b to Z do not make a concomitant large increase from its value of -0.85 at $q = 0$, it follows that the net Z plunges from about $+1.2$ toward a negative value of the order of -1, and that this plunge has begun even for very small q.

4.4. Exact High-Density Calculation

Of course the existence of this cutoff on $Z_{\text{fluc}}(q)$ is based on an actual calculation, and not just the perhaps speculative arguments above. Langreth and Vosko /5/ have numerically evaluated the diagrams of Figure 1 in the high density limit for q in the range $0 \leq q \leq 2k_{\text{FT}}$. The results are shown in Figure 2. One sees clearly that Z has substantially fallen from its $q = 0$ value even for values of q as small as $q \sim k_{\text{FT}}/\sqrt{3}$. For values of $q < k_{\text{FT}}/\sqrt{3}$ one can obtain $Z(q)$ from the formula

$$Z(q) \approx 1.20 + 0.77Q\ln Q - 1.25Q + \ldots \tag{56}$$

where $Q = \sqrt{3}q/k_{\text{FT}}$. This formula fits our numerical results to within 1% for q within the indicated range.

What happens as one increases q even further, beyond what is shown in Fig. 2? For example, one might consider sufficiently high density that the condition $k_{\text{FT}}/\sqrt{3} \ll q \ll 2k_{\text{F}}$ is possible. What would be the value of $Z(q)$ for q in this range? This question was not addressed in Ref. /5/. We have recently shown /42/ analytically that $Z_{\text{ex}}(q)$ slowly approaches the value $-\frac{10}{9}$ in this range. The fact that $-\frac{10}{9}$ is precisely $\frac{10}{7}Z_{\text{Sham}}$ fits in beautifully with

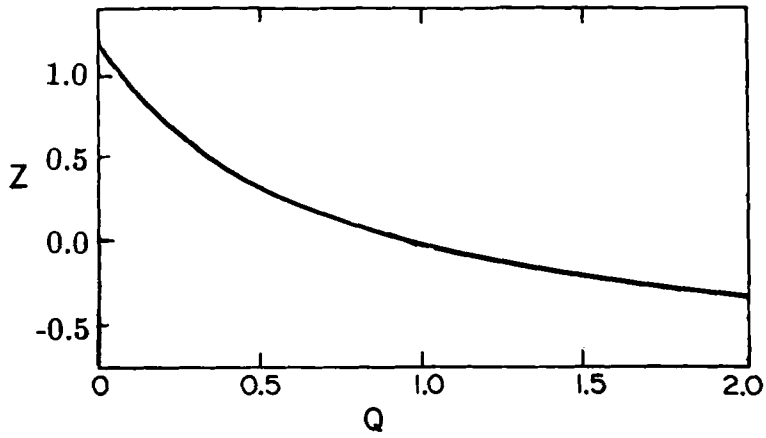


Fig. 2. The Langreth-Vosko calculation of $Z(q)$. The vertical axis represents $Z(q)$ and the horizontal axis represents $Q = q\sqrt{3}/k_{\mathrm{FT}}$. At high density this calculation is exact, and contrasts markedly with the gradient approximation, which consists of assuming that $Z(q)$ is a constant equal to 1.2.

the recently discovered exchange anomalies discussed in the next section. Further discussion will be put off until then.

5. LONG-RANGE ANOMALIES FOR EXCHANGE AND CORRELATION SEPARATELY

5.1. Discovery of the Problem

So far we have been considering the exchange-correlation functional $E_{\mathrm{xc}}[n]$, with exchange and correlation added together and treated on the same footing, rather than trying to write one functional for exchange and another for correlation. On the other hand, the early tradition in density functional theory, beginning with Kohn and Sham /2/, Ma and Brueckner /7/, Sham /10/ and others was to separate the exchange contribution from the correlation contribution. This of course introduces an unphysical long-range (i. e., unscreened) component in each, which results from the long range of the Coulomb potential. It has long since been known that such a separation introduces a spurious singularity in the electronic specific heat /43/ of a uniform system, and it was pointed out by Lang and Sham /44/ and Langreth and Perdew /45/ that such a separation might cause difficulties in density functional theory as applied to extended systems; moreover an inconsistency

with the traditional separation within the context of the local density approximation was pointed out for localized systems /4/. Nevertheless it has not been appreciated until recently that such a separation might also introduce difficulties in the approximate forms of K_{exchange} or $K_{\text{correlation}}$ at $q \sim 0$.

We begin our discussion with the calculation of Sham /10/. He calculated the contribution to $Z(0)$ from the diagrams of Figure 1 a and b with the wiggly lines replaced by bare Coulomb lines [this is the exchange only approximation, mentioned in Ref. /6/]. He used, not the full long-ranged Coulomb interaction, but the limit of Yukawa interaction as the screening parameter k_{screen} went to zero. This result has been duplicated by many others. However, Kleinman /46/, working directly with a Coulomb interaction, obtained a different answer for $Z(0)$ from the same diagrams [in this case, diagrammatic techniques were not used, but the integrals obtained were the same as would have been obtained diagrammatically]. He suggested that there exists a discontinuity/47/ as $k_{\text{screen}} \to 0$.

5.2. Numerical Verification

The first numerical calculation of $\delta\chi(q)$ and hence $Z(q)$ for exchange only was made by Geldart and Taylor /48/. Sham /10/ pointed out that what in our notation are the $\delta\tilde{Z}(q)$'s calculated by Geldart and Taylor in the region $0 < q < \frac{1}{2}k_{\text{F}}$ were "reasonably fit" by a constant equal to $\frac{2}{5}\tilde{Z}_{\text{Sham}}$ [$\frac{2}{5}$ is the ratio of the first two entries in Sham's table I. In our notation, these entries are the values of $\frac{1}{8}\delta\tilde{Z}(q)$]. This means that in this q range

$$\tilde{Z}_{\text{exchange}}(q) \approx \tfrac{2}{5}\tilde{Z}_{\text{Sham}}. \tag{57}$$

According to (44), this translates into the statement that

$$Z_{\text{exchange}}(q) \approx \tfrac{10}{7}Z_{\text{Sham}}. \tag{58}$$

At this time there was no suggestion, however, that there was anything unusual happening at small q.

However, the Geldart–Taylor calculation has been repeated with modern computer power by Antoniweicz and Kleinman /49/, who were able to achieve more accuracy and to push the calculation to smaller q. Based on their results, they conjectured that as q approached zero, that $Z_{\text{exchange}}(q)$ would not turn over smoothly to Z_{Sham} but would rather remain at the value of approximately $\frac{10}{7}Z_{\text{Sham}}$ This was strengthened more recently by the calculations of Kleinman and Lee /50/, who studied the behavior of Z as a function of screening parameter, and by Chevary and Vosko /51/. The latter authors have numerically, but with high precision, evaluated these diagrams

for small q, and obtain an extrapolated value of

$$\tilde{Z}_{\text{exchange}}(q \to 0) = \tfrac{2}{9} = \tfrac{2}{5}\tilde{Z}_{\text{Sham}} \tag{59}$$

to many significant figures. In terms of Z this translates as

$$Z_{\text{exchange}}(q \to 0) = -\tfrac{10}{9} = \tfrac{10}{7}Z_{\text{Sham}}, \tag{60}$$

thus confirming the result anticipated by Antoniweicz and Kleinman.

5.3. Analytic Proof

Still more recently, Langreth and Vosko /42/ have obtained the result (60) *analytically*, by evaluating our result for the graphs a and b of Fig. 1 with the dielectric function set equal to unity. In fact, we have also succeeded in doing the Kleinman–Lee /50/ calculation of $Z_{\text{exchange}}(q)$ for a Yukawa interaction analytically. For $k_{\text{screen}} = 0$ and then $q \to 0$ we find that $Z_{\text{exchange}}(q) \to \frac{10}{7}Z_{\text{Sham}}$, while for $q = 0$ and then $k_{\text{screen}} \to 0$ we find $Z_{\text{exchange}}(q) \to Z_{\text{Sham}}$, thus analytically confirming the discontinuity suggested by the previous numerical work. This apparently arises from the fact that by separating exchange from correlation, one introduces a sensitive dependence on the form of the interparticle interaction at very large distances. This is probably unphysical and it means that attempts to separate exchange and correlation in density functional schemes based on extrapolations from the non-uniform electron gas will have to be made with great care. In particular, one must ensure that the long range anomalies introduced by the process are treated in the same manner in the approximation used for correlation as they are in the approximation made for exchange.

Recently Langreth and Perdew /52/ have reanalyzed their calculation /3/ and showed that the above discontinuity turns up in the form of delta functions at $k = 0$ in the wave-vector decompositions [k can be regarded qualitatively as the k in Fig. 1]. These delta functions are multiplied by ϵ^{-1} where ϵ is the zero frequency dielectric function, which is infinite at $k = 0$. This shows explicitly that these anomalies are *not* present when exchange and correlation are treated together, as we have done in the bulk of this article. However, for correlation alone, the delta functions would be multiplied by $(\epsilon^{-1} - 1)$ and for exchange alone would be multiplied by 1, and would therefore survive.

5.4. Implications for Exchange and Correlation Functionals

At first sight, one might think that the above anomaly is inconsistent with the assumptions made in various calculations such as ours /5/. This is not the case, however, because in our calculation of the diagrams a and b, the Coulomb potential was multiplied by $\epsilon^{-1}(k,\omega)$ and *not* $\epsilon^{-1}(k,\omega)-1$. That is to say, we calculated the exchange and correlation together, and not correlation alone. Nevertheless as q approached zero, we found that $Z(q)$ approached $Z_{\text{Ma-Brueckner}} + Z_{\text{Sham}}$. This is a correct result, whose physics is outlined below. However, as mentioned as a possibility in our letter, our separation into exchange and correlation pieces via the traditional prescription was not correct, because of the long-range anomalies discussed here. Said differently, the value calculated by Ma and Brueckner was *not* the the correlation energy, but rather the *long-range* (or small k) component of the correlation *plus* exchange energies. To it must be added the "non-long-range" component of the exchange energy. The latter is what Sham calculated. In principle, the "non-long-range" part of the correlation energy would also have to be added, but that is of the order of e^4 and hence vanishes in the high density limit. In summary then, the correct values of the exchange and correlation gradient coefficients are

$$Z_{\text{exchange}}(q \to 0) = Z_{\text{Sham}} - \tfrac{1}{3} = \tfrac{10}{7} Z_{\text{Sham}} = -\tfrac{10}{9} \tag{61}$$

and

$$Z_{\text{correlation}}(q \to 0) = Z_{\text{Ma-Brueckner}} + \tfrac{1}{3} = 2.3089. \tag{62}$$

Although the final result for exchange plus correlation is unchanged, the values for exchange and correleation separately are modified due to these long range anomalies. In particular, the contributions of the individual diagrams to $\tilde{Z}$ for $q \to 0$, which were first worked out by Rasolt and Geldart /16/, must now be modified. Table 1 gives the detailed changes necessary.

What about for $k_{\text{FT}}/\sqrt{3} \ll q \ll 2k_{\text{F}}$? Here the fact that an extension of the Langreth-Vosko calculation into this region yields the result $Z(q) = -\frac{10}{9}$ means that it will match on to a good calculation for exchange-only, such as made by Chevary and Vosko, and hence be extended to arbitrary q. This is physical, since at high densities, one expects the correlation contribution to be negligible for $q \gg k_{\text{FT}}$.

In summary: for $q \to 0$ at high density

$$Z(q) \to Z_{\text{Sham}} + Z_{\text{Ma-Brueckner}} = 1.1978, \tag{63}$$

Table 1. Contributions of individual diagrams to $\tilde{Z}(q)$ and $Z(q)$ for $q \to 0$ in the high density limit. In the column labeled "LV", we use the notation a_0 [b_0] for the bare or unscreened part of diagram a [b] of Fig. 1 of this article and also of Ref. /5/. Similarly we use a_1 [b_1] for the screening part. In the column labeled "RG", the letters refer to the diagrams in Fig. 1 of Ref. /16/.

Description	*Diagram*		$\tilde{Z}$ *Values*		Z *Values*	
	RG	*LV*	*Old*	*New*	*Old*	*New*
pure exch.	a	$a_0 + b_0$	5/9	2/9	−7/9	−10/9
scrng of exch.	b	$a_1 + b_1$	−0.0714	0.2619	−0.0714	0.2619
screened exch.	$a + b$	$a + b$	0.4842	0.4842	−0.8491	−0.8491
fluctuation	c	c	2.0470	2.0470	2.0470	2.0470
correlation	$b + c$	$a_1 + b_1 + c$	1.9756	2.3089	1.9756	2.3089
exch.–corr.	$a + b + c$	$a + b + c$	2.5311	2.5311	1.1978	1.1978

and for $k_{\mathrm{FT}}/\sqrt{3} \ll q \ll 2k_{\mathrm{F}}$,

$$Z(q) \to \tfrac{10}{7} Z_{\mathrm{Sham}} = -\tfrac{10}{9}. \tag{64}$$

As q increases from zero, $Z(q)$ falls from its value in (63), rapidly at first, and then more slowly, and if the density is high enough, approaches (64) before decreasing further as q becomes a non-negligible fraction of $2k_{\mathrm{F}}$.

6. CONCLUSIONS AND IMPLICATIONS FOR THE DEVELOPMENT OF NONLOCAL FUNCTIONALS

We have spent the bulk of this article developing the properties of the response function which determines the nonlocal density functional in some limiting cases. We have shown that the nonlocal part of K_{xc}, when expressed through $Z(q)$, is large and *positive* for $q = 0$ (the gradient expansion) but dives to a fairly large *negative* value for very small but finite q. This is a physical feature, which surely has to be taken into account in the development of nonlocal functionals. The simple gradient approximation described earlier does not include this feature, and simply assumes that $Z(q)$ is constant for an appreciable range of q around $q \sim 0$; thus this approximation is not useful for anything but the slowest varying disturbances. The first workable functional for exchange and correlation together that explicitly accounted for this feature was developed by Langreth and Mehl /4/. Perdew's more recent functional /22, 23/ also explicitly accounts for it (at least with respect to exchange and correlation together). These are first steps, and

as all approximate functionals which have been proposed, leave plenty of room for improvement. It is not our intention here to detail the merits or drawbacks of any particular functional, however, but simply to drive home the point that anyone developing a nonlocal functional which purports to describe both slowly and more rapidly varying densities, cannot afford to ignore the simple physical feature discussed here, which the exact functional must have.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation Grant No. DMR-88-01027 and by the Natural Sciences and Engineering Research Council of Canada Grant No. A6294.

REFERENCES

1. P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
2. W. Kohn and L. J. Sham, Phys. Rev. **140**, A1333 (1965).
3. D. C. Langreth and J. P. Perdew, Phys. Rev. B **21**, 5469 (1980).
4. D. C. Langreth and M. J. Mehl, Phys. Rev. B **28**, 1809 (1983).
5. D. C. Langreth and S. H. Vosko, Phys. Rev. Lett. **59**, 497 (1987).
6. It was originally thought that the separation of exchange and correlation was straightforward, with the gradient coefficient for exchange determined by the Hartree-Fock density response function. It has since been shown by a number of authors: L. Kleinman, Phys. Rev. B **10**, 2221 (1974); **12**, 3512 (1975), A. K. Rajagopal and S. Ray, Phys. Rev. B **12**, 3129 (1975), D. J. W. Geldart, M. Rasolt, and C. O. Ambladh, Solid State Commun. **16**, 243 (1975), that the Hartree-Fock gradient expansion for the Coulomb interaction does not exist. However, this problem does not occur in the "exchange only" method of density functional theory as emphasized by V. Sahni, J. Gruenebaum and J. P. Perdew, Phys. Rev. B **26**, 4371 (1982), and it is this definition for exchange that is currently used in density functional theory, which we follow.
7. S.-K. Ma and K. A. Brueckner, Phys. Rev. **165**, 18 (1968).
8. In fact, as discussed in section 5, the high density gradient coefficient for correlation is even larger.

9. F. Herman and J. P. van Dyke and I. B. Ortenburger, Phys. Rev. Lett. **22**, 807 (1969); Int. J. Quantum Chem. **3S**, 827 (1970).

10. L. J. Sham, in *Computational Methods in Band Theory*, edited by P. M. Marcus, J. F. Janak, and A. R. Williams, Plenum, New York, 1971.

11. The coefficient that Sham calculated turns out to be the appropriate one for exchange only, except for the difficulty associated with using a statically screened Coulomb interaction, and then allowing the screening length to become infinite at the end of the calculation. As discussed later in this article, the appropriate coefficient for exchange only turns out to be $\frac{10}{7}$ of the original Sham result; this difference is not sufficiently large to change the conclusion that the theoretical coefficient is significantly smaller than the semi-empirical one.

12. Despite the "10/7" correction to the exchange only coefficient, the Sham coefficient is the correct one to add to the Ma-Brueckner value to obtain the high-density limit of the gradient coefficient for exchange and correlation treated together. However, the correlation coefficient is increased approximately 20%, making the disagreement with experiment even worse. This is discussed in Section 4.

13. D. C. Langreth, in *Many-Body Phenomena at Surfaces*, edited by D. C. Langreth and H. Suhl, page 51, Academic Press, Orlando, 1984.

14. D. C. Langreth, in *Density Matrices and Density Functionals: Proceedings of the A. John Coleman Symposium*, edited by R. Erdahl and V. H. Smith, page 375, Reidel, Dordrecht, 1987.

15. D. J. W. Geldart and M. Rasolt, Phys. Rev. B **13**, 1477 (1976).

16. M. Rasolt and D. J. W. Geldart, Phys. Rev. Lett. **35**, 1234 (1975).

17. P. Bagno, O. Jepsen, and O. Gunnarsson, Phys. Rev. B **40**, 1997 (1989).

18. U. von Barth, in *The Electronic Structure of Complex Systems*, edited by P. Phariseau and W. Temmerman, NATP ASI, Series B: Physica **113**, Plenum, New York, 1984.

19. U. von Barth and A. C. Pedroza, Physica Scripta **32**, 353 (1985).

20. D. C. Langreth and M. J. Mehl, Phys. Rev. Lett. **47**, 446 (1981).

21. C.-D. Hu and D. C. Langreth, Physica Scripta **32**, 391 (1985).

22. J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).

23. J. P. Perdew and Wang Yue, Phys. Rev. B **33**, 8800 (1986), Errata: Phys. Rev. B **34**, 7406 (1986). In subsequent publications the second author's surname (Wang) appears last.

24. J. B. Lagowski and S. H. Vosko, Phys. Rev. A **39**, 4972 (1989).

25. R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

26. We generally follow the notation of Pines and Nozières, Ref. /27/.

27. D. Pines and P. Nozières, *Quantum Liquids*, W. A. Benjamin, Inc., New York, 1966.

28. We follow the original definition of Hohenberg and Kohn, which has been followed in the literature by a number of authors, including ourselves. Another common definition differs by a factor of two.

29. E. Lieb, in *Physics as Natural Philosophy: Essays in Honor of Laszlo Tisza on His 75th Birthday*, edited by A. Shimony and H. Feshbach, page 111, MIT, Cambridge, 1982, Reprinted in Int. J. Quantum Chem. **24**, 243 (1983).

30. M. Levy, Phys. Rev. A **26**, 1200 (1982).

31. D. C. Langreth, Phys. Rev. Lett. **52**, 2317 (1984).

32. See the note added in proof in Ref. /2/.

33. Note that we use units where the system quantization volume, the electronic charge, and $\hbar$ are unity. The quantity $E_{\mathbf{xc}}(n) \equiv n\epsilon_{\mathbf{xc}}(n)$, where $\epsilon_{\mathbf{xc}}(n)$ is the *per particle* exchange-correlation of the uniform system with electronic density n.

34. S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).

35. J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5075 (1981).

36. D. M. Ceperly and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

37. This is the definition actually used by Langreth and Vosko in Ref. /5/. Eq. (4) in Ref. /5/ has a factor of two error, and should be identical to Eq. (32) of the present article [except that here $e^2 = 1$].

38. D. J. W. Geldart and S. H. Vosko, Can. J. Phys. **44**, 2137 (1966).

39. The following two equations for the total $\delta\tilde{Z}$ and Z are unaffected by the long range anomaly discussed in Section 5.

40. D. C. Langreth and J. P. Perdew, Solid State Commun. **31**, 567 (1979).

41. As presented below, the Geldart-Rasolt results for the screened-exchange are still correct, even when the long-range anomalies discussed in Section 5 are accounted for. However, their division of $\tilde{Z}$ betweeen unscreened exchange (5/9) and the screening contribution (−.0714) must be modified because of the recent results.

42. D. C. Langreth and S. H. Vosko, private communication (1989).

43. J. Bardeen, Phys. Rev. **50**, 1098 (1936).

44. N. D. Lang and L. Sham, Solid State Commun. **17**, 581 (1975).

45. D. C. Langreth and J. P. Perdew, Solid State Commun. **17**, 1425 (1975).

46. L. Kleinman, Phys. Rev. B **30**, 2223 (1984).

47. The calculated coefficient of the discontinuity was not consistent with the $\frac{10}{7}Z_{\mathbf{Sham}}$ value, and has been retracted /50/. It is therefore an open question whether the correct value can be obtained by this method.

48. D. J. W. Geldart and R. Taylor, Can. J. Physics **48**, 155 (1970).

49. P. R. Antoniweicz and L. Kleinman, Phys. Rev. B **31**, 6779 (1985).

50. L. Kleinman and S. Lee, Phys. Rev. **37**, 4634 (1988).

51. J. Chevary and S. H. Vosko, Bull. Am. Phys. Soc. **33**, 238 (1988) and University of Toronto Preprint (1989).

52. D. C. Langreth and J. P. Perdew, private communication (1989).

DENSITY-GRADIENT EXPANSIONS

Leonard Kleinman

Department of Physics
The University of Texas
Austin, TX 78712

Viraht Sahni

Department of Physics
Brooklyn College of the City University
of New York
Brooklyn, New York 11210
and
The Graduate School and University Center
of the City University of New York
33 West 42nd Street
New York, New York 10036

This article will be limited to a discussion of the expansions[1,2] of the exchange (x) and correlation (c) energy density functionals. In the first part of this paper recent developments with regard to first principle derivations of the coefficients of these expansions are discussed. The second part deals with an analysis of these expansions based on a study of the corresponding Fermi and Coulomb hole charge distributions.

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

We expand the energy of an electron gas subjected to a weak external potential $v(\mathbf{r})$ whose average may be taken to be zero:

$$E = \int d\mathbf{r} v(\mathbf{r})\rho(\mathbf{r}) + 1/2 \int\int d\mathbf{r}d\mathbf{r}' \tilde{\rho}(\mathbf{r})\tilde{\rho}(\mathbf{r}')/|\mathbf{r} - \mathbf{r}'| + \int d\mathbf{r} g_o[\rho(\mathbf{r})] + 1/2 \int\int d\mathbf{r}d\mathbf{r}' K(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r})\rho(\mathbf{r}'), \quad (1)$$

where $\rho(\mathbf{r})$ is the electronic charge density, $\tilde{\rho}(\mathbf{r})=\rho(\mathbf{r})-n_o$ where n_o is the average charge density which is cancelled by the jellium background and $\int d\mathbf{r} K(\mathbf{r})=0$ because the entire xc and kinetic energies are contained in $\int d\mathbf{r} g_o[\rho(\mathbf{r})]$ when $\rho(\mathbf{r})=n_o$. Note that this expansion differs slightly from that in the original Hohenberg-Kohn (HK) paper.[1] Taking the variation of E with respect to $\rho(\mathbf{r})$ yields

$$v(\mathbf{r}) + \int d\mathbf{r}' \tilde{\rho}(\mathbf{r}')/|\mathbf{r} - \mathbf{r}'| + g_o'[\rho(\mathbf{r})] + \int d\mathbf{r}' K(\mathbf{r} - \mathbf{r}')\tilde{\rho}(\mathbf{r}') - \mu = 0, \quad (2)$$

where the chemical potential μ enters as the Lagrangian multiplier for the conservation of charge condition $\int d\mathbf{r}\rho(\mathbf{r})=N$. Noting that μ can not depend on the sign of $v(\mathbf{r})$, to first order in $v(\mathbf{r})$ and $\tilde{\rho}(\mathbf{r})$, $\mu=\mu_o$; thus equating terms of equal order in $v(\mathbf{r})$, $\rho(\mathbf{r})$, and μ

$$g_o'(n_o) - \mu_o = 0 \quad (3)$$

$$\varphi(\mathbf{r}) + g_o''(n_o)\tilde{\rho}(\mathbf{r}) + \int d\mathbf{r}' K(\mathbf{r} - \mathbf{r}')\tilde{\rho}(\mathbf{r}') = 0, \quad (4)$$

where $\varphi(\mathbf{r})$ represents the external plus Hartree potential (first two terms in Eq.(2)) and we have used $g_o'[\rho(\mathbf{r})]=g_o'(n_o)+g_o''(n_o)\tilde{\rho}(\mathbf{r})$. Writing $v(\mathbf{r})= \Sigma_{\mathbf{K}} v(\mathbf{K})e^{i\mathbf{K}.\mathbf{r}}$, Eq. (1) and (4) become

$$E = \Sigma_{\mathbf{K}} v(\mathbf{K})\rho^*(\mathbf{K}) + \frac{4\pi}{2} \Sigma_{\mathbf{K}} \rho(\mathbf{K})\rho^*(\mathbf{K})/K^2 + g_o(n_o) + \frac{1}{2} \Sigma_{\mathbf{K}}\{g_o''(n_o) + K^2 Z(K)\}\rho(\mathbf{K})\rho^*(\mathbf{K}) \quad (5)$$

$$\varphi(\mathbf{K}) + \{g_o''(n_o) + K^2 Z(K)\}\rho(K) = 0. \tag{6}$$

Here we have used the fact that the zeroth Fourier transform of $K(\mathbf{r})$ vanishes and that $K(\mathbf{K})$ cannot depend on the direction of $\mathbf{K}$, to write $K(\mathbf{K})=K^2Z(K)$. Now the irreducible density response function is defined by

$$\rho(\mathbf{K}) = \bar{\chi}(\mathbf{K})\varphi(\mathbf{K}) \tag{7}$$

so that

$$\{g_o''(n_o) + K^2 Z(K)\} = -1/\bar{\chi}(K) \tag{8}$$

and therefore, separating g_o and Z into their kinetic and xc components

$$\{g_{oxc}''(n_o) + K^2 Z_{xc}(K)\} = 1/\bar{\chi}_{RPA}(K) - 1/\bar{\chi}(K). \tag{9}$$

If $\tilde{\rho}(\mathbf{r})$ is not only small but also is slowly varying, the last term in Eq. (1) can be replaced by

$$E_{grad} = 1/2 \int g_2(\rho)|\nabla\rho(\mathbf{r})|^2 d\mathbf{r}. \tag{10}$$

Noting that $\rho(\mathbf{r})=n_o+\Sigma_{\mathbf{K}}\rho(\mathbf{K})e^{i\mathbf{K}.\mathbf{r}}$, it is easy to see that the gradient expansion coefficient $g_2(\rho)$ is identical to Z (K→0). Expanding $\bar{\chi}(K)=\bar{\chi}^{(0)}+\bar{\chi}^{(2)}K^2$ yields

$$g_o''(n_o) = -1/\bar{\chi}^{(o)}, \tag{11}$$

which yields the well known compressibility sum rule,[3] $\bar{\chi}^{(0)} = -n_o^2 K$, where K is the compressibility and

$$g_{2xc} = Z_{xc}(K\to 0) = \bar{\chi}^{(2)}/\bar{\chi}^{(o)2} - \bar{\chi}^{(2)}_{RPA} / \bar{\chi}^{(0)2}_{RPA}. \tag{12}$$

Only one year after HK, Kohn and Sham (KS)[2] showed that $\rho(\mathbf{r})$ was obtainable from a set of one-electron eigenfunctions of a Schroedinger equation:

$$\rho(\mathbf{r}) = \Sigma_i |\psi_i(\mathbf{r})|^2 \tag{13}$$

$$[-\nabla^2 + \varphi(\mathbf{r}) + v_{xc}(\mathbf{r}) - \varepsilon_i]\psi_i(\mathbf{r}) = 0, \tag{14}$$

where $v_{xc}(\mathbf{r})=\delta E[\rho(\mathbf{r})]/\delta\rho(\mathbf{r})$. In their original work, KS pointed out that ε_i and ψ_i have no physical meaning and that only $\rho(\mathbf{r})$ and the total energy E obtained from it are meaningful. Recent work,[4,5] however, has shown that the highest occupied eigenvalue ε_{max} of the KS equation does have the physical interpretation of being the negative of the ionization potential.

It is interesting to note that the first calculation of a gradient expansion coefficient was performed by Ma and Brueckner (MB)[6] for

$$g_{2c} = \bar{\chi}^{(2)}/\bar{\chi}^{(0)2} - \bar{\chi}_x^{(2)}/\bar{\chi}_x^{(0)2} \tag{15}$$

rather than for g_x or g_{xc}. Perhaps they anticipated the problems with g_x that are only now being resolved. The first attempt to obtain the gradient expansion for exchange was a parameterization scheme of Herman, Ortenburger and Van Dyke[7] (HOvD) who wrote

$$E_x = \frac{3\alpha}{4}\int \rho V_{xs}(\rho)d\mathbf{r} + \frac{3\beta}{2}\int \rho G(\rho)V_{xs}(\rho)d\mathbf{r} \tag{16}$$

where

$$G = (1/\rho^{2/3})[(4/3)(\nabla\rho/\rho)^2 - (2\nabla^2\rho/\rho)] \tag{17}$$

is a dimensionless function and V_{xs} is the Slater (local density approximation) exchange potential (in Ry.),

$$V_{xs} = -3\,[3\rho(\mathbf{r})/\pi]^{1/3}. \tag{18}$$

With $\alpha=2/3$ the first term is $\int d\mathbf{r} g_{ox}[\rho(\mathbf{r})]$ when the functional form of $g_{ox}(\rho)$ is obtained from the exchange energy of the free

electron gas. The second term was chosen to have the same $\rho^{4/3}$ dependence as the first, and after an integration by parts to eliminate the $\nabla^2\rho$ term, becomes identical to Eq. (10) if $g_2(\rho) = -3\beta(3/\pi)^{1/3}\rho^{-4/3}$. Taking the functional derivative of Eq. 16 leads to the exchange potential

$$v_x(\rho) = \alpha V_{xs}(\rho) + \beta G(\rho) V_{xs}(\rho). \qquad (19)$$

HOvD then varied α and β to obtain eigenfunctions which minimized the expectation value of the Hartree Fock Hamiltonian for Ar, Cu^+, Kr and Xe. In each case the optimal value of α was 2/3 and the β were close to 0.005. The atomic energies obtained in this manner were very close to Hartree Fock energies. For example, for Ar and Kr the HOvD total energies are -526.81a.u. and -2752.04a.u. respectively. The corresponding Hartree-Fock results[8] for these atoms are -526.82a.u. and -2752.06a.u. It is interesting to note that the HOvD results are precisely the same as those obtained by Talman and co-workers[9] within the framework of the optimized potential method (OPM). The OPM is a variational scheme (i.e. no gradient approximation) that obtains a local potential whose eigenfunctions mimimize the Hartree-Fock energy. Thus, two self-consistently determined local exchange potentials lead to the same minimum value of the Hartree-Fock Hamiltonian. The HOvD highest occupied eigenenergy for Kr, however, is -0.62Ry, whereas the Talman et al and Hartree-Fock values are -1.00Ry and -1.05Ry respectively. The explanation of why a gradient expansion approximation cannot lead to accurate highest occupied eigenenergies is discussed later in this article. For Xe, the HOvD total energy is -7232.13a.u. and is in fact superior to the Talman result which is -7232.12 a.u. (Presumably, there is a computational error in one of the two papers). The Hartree-Fock result is -7232.14a.u. HOvD then fixed α at 2/3 and varied β to minimize the total energy obtained from the

density functional Hamiltonian for Kr. This gave -2752.03a.u. for β=0.00314. Were it sufficient to stop the expansion at the gradient term, there would have been only one β and for that β the energy obtained from the density functional and HF Hamiltonians would have been identical. Nevertheless, that energy would still lie slightly above the HF energy.

The first attempt to obtain g_{2x} from first principles is due to Sham[10] who argued that since exchange is first order in e^2, he need only obtain the $\bar{\chi}^{(2)}$ and $\bar{\chi}^{(0)}$ in Eq. (12) to first order in e^2. This he did by writing the HF exchange plus Coulomb potential in terms of φ, the Coulomb potential

$$v(\mathbf{k} + \mathbf{K}, \mathbf{k}) = \tilde{\Lambda}(\mathbf{k} + \mathbf{K}, \mathbf{k})\varphi(\mathbf{K}). \tag{20}$$

Applying perturbation theory in the HF approximation gave

$$\tilde{\Lambda}(\mathbf{k} + \mathbf{K}, \mathbf{k}) = 1 - (1/2\pi)^3 \int d\mathbf{k}'\Theta(\mathbf{k}' + \mathbf{K}, \mathbf{k}')u(\mathbf{k} - \mathbf{k}') \tilde{\Lambda}(\mathbf{k}' + \mathbf{K}, \mathbf{k}'), \tag{21}$$

where

$$u(q) = 4\pi e^2/q^2 \tag{22}$$

$$\Theta(\mathbf{k}' + \mathbf{K}, \mathbf{k}') = (f_{\mathbf{k}'+\mathbf{K}} - f_{\mathbf{k}'})/[\varepsilon(\mathbf{k}' + \mathbf{K}) - \varepsilon(\mathbf{k}')], \tag{23}$$

$f_{\mathbf{k}}$ is the Fermi function and $\varepsilon(\mathbf{k})$ is the kinetic plus HF exchange energy of an electron of wave vector $\mathbf{k}$ in a free electron gas. From Eq. (20) and (7) one obtains

$$\bar{\chi} = 2/(2\pi)^3 \int d\mathbf{k}\Theta(\mathbf{k} + \mathbf{K}, \mathbf{k})\tilde{\Lambda}(\mathbf{k} + \mathbf{K}, \mathbf{k}). \tag{24}$$

Expanding $\tilde{\Lambda}$ in a power series in e^2 and equating terms of the same order in e^2 the integral equation (21) was solved to order K^2. In order to make the integral tractable, q^2 in (22) was replaced by $q^2+\lambda^2$ and the $\lambda\to 0$ limit taken as the last step in the calculation. Sham obtained (in Hartree atomic units)

$$g_{2x}^{Sham} = -[7/216\pi(3\pi^2)^{1/3}]\rho^{-4/3}. \quad (25)$$

If one inserts (17) in (16), integrates by parts to put (16) in the form of (10), and converts Rydbergs to Hartrees, one obtains $g_{2x}^{HOvD}=-3\beta(3/\pi)^{1/3}\rho^{-4/3}$ which for $\beta=.00314$ is 2.78 times larger than Sham's result. In Hartree-Fock theory (HFT), the exchange energy contains terms to all orders in the electron-electron interaction, because the orbitals depend explicity on e^2. In the evaluation of the HFT coefficient g_{2x}^{HFT} one must thus sum to all orders in e^2. Such a calculation has been performed for screened Coulomb interaction by Kleinman[11] and others[12], and the resulting expression is

$$g_{2x}^{HFT} = -\frac{\pi}{24(3\pi^2)^{4/3}}[\tfrac{1}{6}(5 - 2\pi k_F + G + D)]\rho^{-4/3}$$

$$G = (1-2\gamma^2)\ln(1+4/\gamma^2) - (64-4\gamma^2-3\gamma^4)/(4+\gamma^2)^2$$

$$D = -E/F$$

$$E = 2[1 + \pi k_F + \gamma^2/(4 + \gamma^2) - (\gamma^2/2)\ln(1 + 4/\gamma^2)]^2$$

$$F = 1 - \pi k_F - (1/2)(1 + \gamma^2/2)\ln(1 + 4/\gamma^2)$$

$$\gamma = \lambda/k_F. \quad (26)$$

This expression diverges[11-13] as $\ln \lambda$ for $\lambda\to 0$. A perturbative analysis[12] shows that it is terms of $O(e^4)$ and higher order that are singular. As a consequence, a gradient expansion for the exchange energy in HFT does not exist. Kleinman[11] pointed out that since the kinetic energy in KS theory is that of KS eigenfunctions, the exchange term contains not only the Fock exchange energy but also the difference between the kinetic energy of the HF and KS eigenfunctions. Thus, it is not of dimensionality e^2. Sahni et al,[14] however, noted that since KS eigenfunctions depend only on the electron density and not otherwise on e, the density gradient expansion of the

density-functional theory exchange energy is purely of $O(e^2)$ and therefore well defined. Thus, in fact what Sham was actually calculating is the Fock energy of Kohn-Sham eigenfunctions, which is of $O(e^2)$. Subsequently, Mohammed and Sahni[15] calculated the density-functional theory (DFT) gradient coefficient for screened-Coulomb interaction to be

$$g_{2x}^{DFT} = -\frac{\pi}{24(3\pi^2)^{4/3}}\,[2-(3\gamma^2/4)\ln(1+4/\gamma^2) -(40-6\gamma^2-3\gamma^4)/3(4+\gamma^2)^2]\rho^{-4/3}. \tag{27}$$

Note that within density-functional theory, both the local density[16] and gradient expansion coefficients are universal functions of the parameter γ whereas the Harteee-Fock theory gradient coefficient is not. Since the total energy has to be independent of basis set, the diference between the HF and KS exchange energies becomes part of the KS correlation energy. Note also that the KS eigenfunctions are much easier to work with; since the KS exchange is wave vector independent, it cancels out of energy denominators as in Eq. (23).

At least three groups[17-19] have calculated g_{2xc} and found it equal to the sum of Sham's g_{2x} and MB's[6] g_{2c}. MB made a diagrammatic expansion of Eq.(15) which they evaluated in the high density limit. Geldart and Rasolt[19] added the (lowest order in e^2) exchange diagrams which MB had subtracted out and did not take the high density limit. Langreth and Perdew (LP)[17] write the xc energy in a wave vector decomposition of the density-density correlation function $S_\alpha(\mathbf{k})=(N/\Omega)\int S_\alpha(\mathbf{r},\mathbf{r}')\exp[i\mathbf{k}\cdot(\mathbf{r}'-\mathbf{r})]d\mathbf{r}d\mathbf{r}'$:

$$E_{xc} = (1/2\pi)^3 \int d\mathbf{k} E_{xc}(\mathbf{k}). \tag{28}$$

where

$$E_{xc}(\mathbf{k}) = (2\pi e^2/k^2)\int_0^1 d\alpha N[S_\alpha(\mathbf{k})-1] \tag{29}$$

and the α integration represents a turning on of the coupling constant e^2 while keeping $\rho(\mathbf{r})$ at its α=1 value. From the fluctuation dissipation theorem

$$\frac{N}{\Omega} S_\alpha(\mathbf{r},\mathbf{r}') = \frac{1}{2\pi i} \int_C d\omega\, \chi(\mathbf{r},\mathbf{r}',\omega), \tag{30}$$

where the contour encloses the positive real axis and χ is defined by

$$\rho(\mathbf{r},\omega) = \int d\mathbf{r}'\, \chi(\mathbf{r},\mathbf{r}',\omega) v(\mathbf{r}',\omega) \tag{31}$$

or

$$\rho(\mathbf{K},\omega) = \chi(\mathbf{K},\omega) v(\mathbf{K},\omega), \tag{32}$$

where $v(\mathbf{K}, \omega)$ is a time dependent external potential. Comparing (32) with (7), where $\varphi(\mathbf{K})$ is the external plus screening potential, results in

$$\chi = \bar{\chi}(1-V\bar{\chi})^{-1}, \tag{33}$$

where $V=\alpha e^2|\mathbf{r}-\mathbf{r}'|^{-1}$ or $4\pi\alpha e^2/K^2$. If one uses χ_{RPA} in (30) then (29) yields exchange plus the contribution of a restricted class of diagrams to the correlation energy which is exact in the high density limit. LP make the gradient expansion for the wave vector decomposition,

$$E_{xc}(\mathbf{k}) = \int d\mathbf{r}[A_{\mathbf{k}}^{xc}(\rho(\mathbf{r})) + \nabla\rho(\mathbf{r})\cdot\overleftrightarrow{B}_{\mathbf{k}}^{xc}(\rho(\mathbf{r}))\cdot\nabla\rho(\mathbf{r})], \tag{34}$$

take $\bar{\chi}$ to be $\bar{\chi}_{RPA}$ which is given by

$$\bar{\chi}_{RPA}(\mathbf{r},\mathbf{r}',\omega) = 2\,\Sigma_{ij}[(f_i - f_j)/(\omega + i\eta + \varepsilon_i - \varepsilon_j)] \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}'), \tag{35}$$

where the ψ_i are one-electron eigenfunctions of the KS equation (14) and f_i is the Fermi function. Then they expand

$$\bar{\chi}_{RPA} = \bar{\chi}_o + \bar{\chi}_1 + \bar{\chi}_2 \tag{36}$$

where

$$\bar{\chi}_o(\mathbf{K},\omega) = \frac{2}{\Omega} \Sigma_{\mathbf{k}} \, (f_{\mathbf{k}} - f_{\mathbf{k}+\mathbf{K}})/(\omega + i\eta + \varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{K}}) \tag{37}$$

is the usual Lindhard susceptibility of a free electron gas with $\varepsilon_k=(1/2)k^2$ and

$$\bar{\chi}_1 = \int d\mathbf{r}_1 \delta\bar{\chi}_{RPA}(\mathbf{r},\mathbf{r}',\omega)/\delta v_{KS}(\mathbf{r}_1) \Big|_{v_{KS}=0} v_{KS}(\mathbf{r}_1) \tag{38}$$

$$\bar{\chi}_2 = (1/2) \int\int d\mathbf{r}_1 d\mathbf{r}_2 \, \delta\bar{\chi}_{RPA}(\mathbf{r},\mathbf{r}',\omega)/\delta v_{KS}(\mathbf{r}_1)\delta v_{KS}(\mathbf{r}_2) \Big|_{v_{KS}=0} v_{KS}(\mathbf{r}_1)v_{KS}(\mathbf{r}_2). \tag{39}$$

Here $v_{KS}(\mathbf{r})=\varphi(\mathbf{r})+v_{xc}(\mathbf{r})$ in Eq. (14) and is eliminated in favor of $\nabla\rho(\mathbf{r})$ by using

$$\rho(\mathbf{K}) = \bar{\chi}_o(\mathbf{K},0)v_{KS}(\mathbf{K}). \tag{40}$$

Note that we have used $\mathbf{k}$ to represent two entirely different things: the wave vector of a single free electron as well as that of the Fourier decompostion of $S(\mathbf{r},\mathbf{r}')$. $\mathbf{K}$ has been used for the wave vector of the external potential and the charge density and potentials it induces. After considerable algebra LP arrive at their final result, which in the high density limit is[20]

$$g_{2x} = -3.333 \times 10^{-3}\rho^{-4/3} \tag{41}$$

$$g_{2xc} = 5.136 \times 10^{-3}\rho^{-4/3}. \tag{42}$$

Here g_{2x} is identical with Eq.(25) and g_{2xc}-g_{2x} is equal to MB's g_{2c}. We can summarize the LP derivation by saying that they obtained the dependence on the gradient of the charge density of the wave vector decomposition of an *exchange-correlation hole calculated from KS eigenfunctions.*

Eq. (41) and (42) were held to be true by the entire density functional physics community until 1984 when Kleinman[21] calculated the exchange energy of KS eigenfunctions

$$\psi_{\mathbf{k}} = (1/\sqrt{\Omega})\{(1 - \alpha_{\mathbf{k}}^2 - \beta_{\mathbf{k}}^2)^{1/2}\exp(i\mathbf{k}\cdot\mathbf{r}) + \alpha_{\mathbf{k}}\exp[i(\mathbf{k} + \mathbf{K})\cdot\mathbf{r}] + \beta_{\mathbf{k}}\exp[- i(\mathbf{k}-\mathbf{K})\cdot\mathbf{r}]\} \quad (43)$$

where

$$\alpha_{\mathbf{k}} = v_{KS}(\mathbf{K})/(K^2/2 + \mathbf{k}\cdot\mathbf{K}),\ \beta_{\mathbf{k}} = v_{KS}(\mathbf{K})/(K^2/2 - \mathbf{k}\cdot\mathbf{K}) \quad (44)$$

and only the $\pm\mathbf{K}$ Fourier components of $v_{KS}(\mathbf{r})$ are considered. The exchange energy of a pair of electrons

$$-\int\int \psi_{\mathbf{k}}(\mathbf{r}_1)\psi^*_{\mathbf{k}'}(\mathbf{r}_1)r_{12}^{-1}\psi^*_{\mathbf{k}}(\mathbf{r}_2)\psi_{\mathbf{k}'}(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \quad (45)$$

was evaluated to obtain

$$\begin{aligned} E_x(\mathbf{k},\mathbf{k}') = -(4\pi/\Omega)[&\{1 - (\alpha_{\mathbf{k}} - \alpha_{\mathbf{k}'})^2 \\ &- (\beta_{\mathbf{k}} - \beta_{\mathbf{k}'})^2\}/(\mathbf{k} - \mathbf{k}')^2 \\ &+ (\alpha_{\mathbf{k}} + \beta_{\mathbf{k}'})^2/(\mathbf{k} - \mathbf{k}' + \mathbf{K})^2 \\ &+ (\alpha_{\mathbf{k}'} + \beta_{\mathbf{k}})^2/(\mathbf{k} - \mathbf{k}' - \mathbf{K})^2] \ . \end{aligned} \quad (46)$$

The total exchange energy per unit volume is

$$E_x = (1/\Omega)[\Omega^2/(2\pi)^6]\int\int E_x(\mathbf{k},\mathbf{k}')f_{\mathbf{k}}f_{\mathbf{k}'}d\mathbf{k}d\mathbf{k}' . \quad (47)$$

Noting that $\beta_{\mathbf{k}+\mathbf{K}}$=-$\alpha_{\mathbf{k}}$ and substituting (44) and (46) into (47) resulted in

$$E_x = -2/(2\pi)^5[\int\int d\mathbf{k}d\mathbf{k}' f_{\mathbf{k}} f_{\mathbf{k}'}/|\mathbf{k} - \mathbf{k}'|^2 + v_{KS}^2(\mathbf{K}) \int\int d\mathbf{k}d\mathbf{k}' \{2/|\mathbf{k} - \mathbf{k}'|^2\} \{1 - (K^2 + 2\mathbf{k}'\cdot\mathbf{K})/(K^2 + 2\mathbf{k}\cdot\mathbf{K})\}\theta_{\mathbf{k}}(\mathbf{K})\theta_{\mathbf{k}'}(\mathbf{K})] \quad (48)$$

where

$$\theta_{\mathbf{k}}(\mathbf{K}) = (f_{\mathbf{k}+\mathbf{K}} - f_{\mathbf{k}})\{(1/2)K^2 + \mathbf{k}\cdot\mathbf{K}\} \approx f'_{\mathbf{k}} + (1/4)(K^2 + 2\mathbf{k}\cdot\mathbf{K})f''_{\mathbf{k}} + (1/24)(K^2 + 2\mathbf{k}\cdot\mathbf{K})^2 f'''_{\mathbf{k}} \quad (49)$$

is expanded to second order in K. The first integral in (48) is just the exchange energy of the free electron gas. The second was evaluated using $f'_{\mathbf{k}}=-\delta(E-E_F)$, $f''_{\mathbf{k}}=-d\delta(E-E_F)/dE$, $f'''_{\mathbf{k}}=-d^2\delta(E-E_F)/dE^2$, $d\mathbf{k}=kd\Omega dE$ and integrating by parts over $dEdE'$ and performing tedious integrals over the solid angles $d\Omega d\Omega'$. The result is, after using Eq. (40) and (37) to eliminate the v_{KS}

$$-E_x = (3/4)(3/\pi)^{1/3} n_o^{4/3} + (1/2)[(1/3)(3/\pi)^{1/3} n_o^{-2/3} + (n_o^{-4/3}K^2)/27\pi(3\pi^2)^{1/3}][|\rho(\mathbf{K})|^2 + |\rho(-\mathbf{K})|^2]. \quad (50)$$

The first term is $g(n_o)$ of Eq. (5) and the last one is $\Sigma_{\mathbf{K}}\{g''_o(n_o) + K^2 Z(0)\}\rho(K)\rho^*(K)$. Comparing the factor of 1/27 with the 7/216 in Eq. (25) we see Z(0) is 8/7 that of Sham.[10] This was shown to be a consequence of Sham using screened exchange. It was argued that the $\lambda\to 0$ limit must be taken before the $K\to 0$ limit in Z(K). Kleinman's factor of 8/7 turned out to be incorrect as he and Antoniewicz(AK)[22] discovered when they evaluated Eq. (48) for all **K**. His mistake was in assuming that the denominators in some 0/0 terms would become finite for nonzero temperatures when $f'_{\mathbf{k}} \neq -\delta(E-E_F)$.

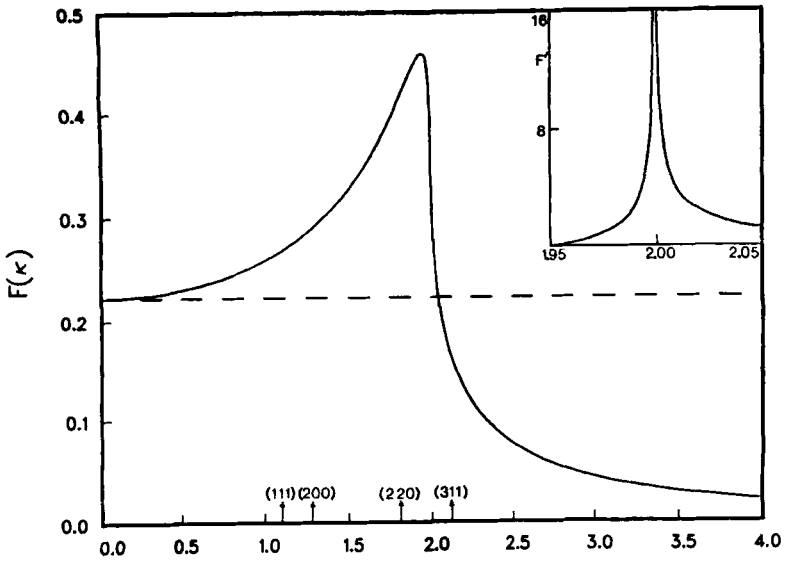


Fig. 1. The dimensionless quantity $F(\kappa)$ versus $\kappa = K/k_F$. Values of κ equivalent to the (111), (200), (220), and (311) reciprocal-lattice vectors in zinc blend are indicated. The horizontal line is the LDA result $F(\kappa) = 2/9$. The inset shows $dF/d\kappa$ in the region it becomes infinite. (From Ref. 22).

AK were able to integrate (48) over $d\theta dk d\theta' dk'$ analytically and numerically over $dk_z dk'_z$. $F(\kappa)$ is plotted in Fig. 1 where $\kappa = K/k_F$ and $3/4(3/\pi)^{1/3} n_o^{-2/3} F(\kappa) = 1/2\{g''_{ox}(n_o)+\kappa^2 Z_x(\kappa)\}$. The dashed line is the local density approximation (LDA) i.e. $F(0) = 2/9$. The four lowest reciprocal lattice vectors of the zinc blende lattice are indicated in Fig. 1. We see that the gradient expansion (i.e. taking $Z(\kappa)$ to be $Z(\kappa \to 0)$) might be used for the first two reciprocal lattice vectors but gives complete nonsense for the rest. For something like Ca with 2 rather than 8 electrons per unit cell, the K_{ijk}/k_F lie a factor of $4^{1/3}$ higher on the κ axis and the gradient expansion is completely useless. Noting

that $n_o^{-2/3}K^2 = (3\pi^2)^{2/3}\kappa^2$ we see that Eq. (50) yields $F(\kappa\to 0) = 2/9 + (2/81)\kappa^2$. Therefore AK plotted $[F(\kappa) - 2/9]/(2/81)\kappa^2$ and found for small κ that the ratio was about 5/4. Because numerical instabilities set in for very small κ, the exact value of the ratio could not be obtained. Thus, according to AK the correct g_{2x} is about 5/4 x 8/7 = 10/7 times as large as Sham's value, Eq.(25). Many of the experts remained unconvinced at this point. Langreth and Vosko[23] wrote that one can use Sham's method to resolve any K~0 anomalies, since such anomalies, if they exist, must cancel out of $g_{2x}+g_{2c}$. Perdew[24] then suggested that if the AK calculation were repeated for screened exchange and the analytic results of Mohammed and Sahni[15] obtained, the doubters ought to be convinced. In Fig. 2 is plotted the result of such a calculation[25] for $\zeta = Z_x(\kappa,\kappa_s)/Z_x^{Sham}$ for several values of κ_s, the screening wave vector, as a function of κ. Even with finite κ_s, numerical instabilities occured for κ=0 so what this figure proves is that the numerically calculated curves for finite κ go smoothly to the analytic result at κ=0 for any non-zero value of κ_s. Note for κ=0, $\kappa_s\to 0$ that $\zeta \to 1$ but for κ_s=0, $\kappa\to 0$ that $\zeta \to 10/7$. Because one-electron Fock exchange is well defined and unscreened, it was concluded

$$g_{2x} \approx \frac{10}{7} g_{2x}^{Sham} \tag{51}$$

is the correct result, but it still was not clear which g_{2x} should be added to MB's g_{2c} to obtain the correct g_{2xc}. Finally, Chevary and Vosko[26] repeated the AK calculation but were more clever in their numerical integration and thus were able to get to smaller κ. They found Eq. (51) to be exact to four figures. It is interesting to note that Sham had inserted Geldart and Taylor's[27] numerically evaluated $\bar{\chi}$ into Eq.(9) and found a $Z_x(\kappa)$ for small κ which was about 10/7 times his own

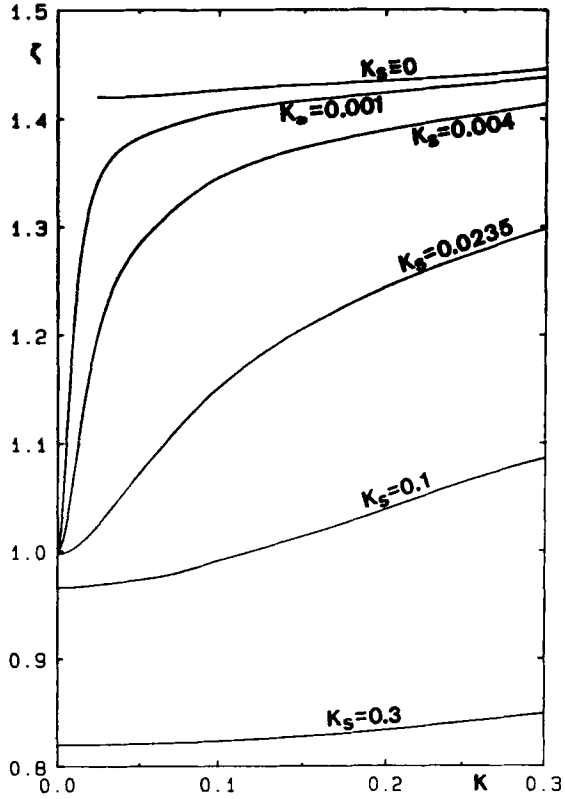


Fig. 2. Plot of $\zeta = Z_x(\kappa,\kappa_s)/Z_x^{Sham}$ as a function of $\kappa = K/k_F$ for various values of κ_s. (From Ref. 25).

Z(0) but did not attach any significance to it. Because $Z_x(\kappa\to0)$ appears to be a rational fraction times Z_x^{Sham}, it seems that if one were clever enough one could obtain $Z_x(\kappa\to0)$ analytically. However $Z_x(\kappa=0)$ does not exist and $Z_x(\kappa\to0)$ cannot be obtained by expanding about $\kappa=0$. The failure of $Z_x(\kappa=0)$ to exist can be thought of in at least three ways: 1) Coulomb integrals are conditionally convergent. Therefore, one can obtain any result one wishes from an infinite wavelength

perturbation. 2) Only $\kappa^2 Z(\kappa)$ is defined. Thus for $\kappa=0$, Z may take any value less infinite than κ^{-2}. 3) The external potential $v(\mathbf{K})$ may be thought of as arising from an external charge density $\rho_{ext}(\mathbf{K})$. Since charge neutrality requires the constant background of charge density to be fixed at n_o, $\rho_{ext}(K=0)$ cannot exist.

Kleinman and Tamura (KT)[28] then reexamined the MB[6] calculation to try to determine which g_{2x} should be added to g_{2c}^{MB}. They first noted that it was remarkable that the lowest order contributions to g_{2c}^{MB} are of order e^2 rather than e^4 as for g_{oc}. MB had given a heuristic argument for this. They assumed the correlation energy is the difference between screened and unscreened exchange interactions. This, of course, is not correct since the screened exchange hole contains no net charge. There is an exchange correlation hole which may be considered to consist of a screened exchange hole plus a Coulomb hole. The exchange hole is due to the antisymmetry of the wave function; the exchange-correlation hole differs from the exchange hole due to electron-electron interactions. Therefore the correlation energy which is the difference in the energy between the two holes must be second and higher order in e^2. Perdew,[29] stimulated by an early version of KT, pointed out that the gradient expansion for correlation is valid only when

$$|\nabla\rho|/\rho << 2k_{FT} \tag{52}$$

$$|\nabla_\mu\nabla_\nu\rho|/|\nabla\rho| << 2k_{FT} \tag{53}$$

where

$$k_{FT} = e\,(4k_F/\pi)^{1/2}. \tag{54}$$

Thus as $e \to 0$ the gradient contribution to the correlation energy does vanish as e^4 in spite of the fact that the coefficient of this term is proportional to e^2.

One of the terms in the MB analysis is

$$b'' = (e^2/k_F^2) \int_0^\infty dq/q \int_{-\infty}^\infty dy(1/\varepsilon - 1)g(q/k_F, y)/2\pi \qquad (55)$$

where ε is the dielectric constant,

$$\varepsilon = 1 + (4e^2 k_F/\pi q^2)R(q/k_F, y), \qquad (56)$$

$y=\omega/iqk_F$ and g and R are dimensionless functions. An expansion of $1/\varepsilon - 1$ is first and higher order in e^2 which would appear to make b" second order in e^2. However MB found a first order contribution which arose at q=0. That is, if the q integration in (55) is broken up into two parts, one between 0 and δ and the other between δ and ∞ and the limit $\delta \to 0$ taken, the entire contribution arises from the 0 to δ integration. MB never considered the possibility that this contribution might vanish if the K=0 limit were taken after rather than before the integration was performed. KT asserted that the e^2 dependence of g_{2c}^{MB} was proven only for K=0, i.e. only when the $|\nabla \rho|^2$ for which it is the coefficient, is identically zero. The $1/\varepsilon$ in (55) arises from screened exchange diagrams and the -1 from subtracting off the corresponding unscreened exchange diagrams. MB took its contribution to be zero in spite of the infinity in the q integration because $\int_{-\infty}^\infty g(q/k_F, y)dy=0$. KT recast this term in a form that looked infinite but in fact this term is actually the indeterminant part of $Z_x(0)$. KT therefore concluded that the MB calculation was not correct but since the same result was obtained by several other workers they allowed for the possibility that by a fortuitous cancellatiion of errors MB had obtained the correct result.

Perdew[29] noted that the gradient contribution to the average exchange hole contains an undamped $\cos(2k_F R)$

oscillation at large R. This results in a delta function $\delta(k/2K_F)$ which must be added to LP's $Z_x(k/2k_F)$, their Eq. (3.52) and (3.55), but whose coefficient cannot be determined by the LP procedure. It is almost certain that this term will account for the difference between g_{2x}^{Sham} and $(10/7)g_{2x}^{Sham}$. What is unresolved at present is whether or not there is a $\delta(k/2k_F)$ with equal coefficient but opposite sign to be added to $Z_c(k/2k_F)$. If so, then the error in g_{2x}^{Sham} is cancelled by the error in g_{2c}^{MB} and $g_{2xc} = g_{2x}^{Sham} + g_{2c}^{MB}$, but if not then (in the high density limit) it is quite likely that

$$g_{2xc} = \frac{10}{7} g_{2x}^{Sham} + g_{2c}^{MB}. \tag{57}$$

The gradient expansion is only useful if Z(K) in Eq. (5) can be approximated by a constant over a large range of K. From Fig. 1 we see that this can be done for Z_x for $K < 1.5k_F$ but $Z_c((2/3)k_F) \approx 0.22Z_c(0)$ when $r_s = 2$ according to Fig. 3 of Ref. 23. Thus the gradient expansion for correlation is never valid. It is possible to use $Z_{xc}(K)$ directly without ever making the gradient expansion. $Z_x(K)$ is given <u>exactly</u> by Fig. 1 and some progress has been made in obtaining $Z_c(K)$. The high density results of Ref. 23 are approximately correct at metallic densities only for very small K and therefore need to be extended to metallic density for all K. This involves not only a more exact evaluation of the diagrams currently being evaluated but also the inclusion of many more higher order diagrams.

Another way to analyze[30] the gradient expansion approximation is in terms of the Fermi-Coulomb hole charge distribution of an electron since the exchange-correlation energy[31-33] is the Coulomb interaction energy between an electron and its Fermi-Coulomb hole. In Kohn-Sham theory, all the many-body effects are incorporated in this charge distribution since the kinetic energy is treated as that of a system of non-interacting electrons. Thus,

$$E_{xc}[\rho] = 1/2 \int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \; \rho(\mathbf{r})\rho_{xc}(\mathbf{r},\mathbf{r}')/|\mathbf{r} - \mathbf{r}'|, \tag{58}$$

where the Fermi-Coulomb hole distribution $\rho_{xc}(\mathbf{r},\mathbf{r}')$ at $\mathbf{r}'$ for an electron at $\mathbf{r}$ is given as

$$\rho_{xc}(\mathbf{r},\mathbf{r}') = \rho(\mathbf{r}') \int_0^1 \mathrm{d}\lambda \; [g_\lambda(\mathbf{r},\mathbf{r}') - 1], \tag{59}$$

and where $g_\lambda(\mathbf{r},\mathbf{r}')$ is the pair-correlation function of a system with density $\rho(\mathbf{r})$ and coupling constant λ. The depletion of charge about each electron is expressed by the charge conservation sum rule

$$\int \rho_{xc}(\mathbf{r},\mathbf{r}')\mathrm{d}\mathbf{r}' = 1. \tag{60}$$

The exact structure of the Fermi-Coulomb hole is, of course, unknown. However, in the exchange-only approximation[14,34] in which only the effects of Pauli correlation are assumed in the wavefunction, the Fermi hole $\rho_x(\mathbf{r},\mathbf{r}')$ is known explicitly in terms of the orbitals which generate the density and is defined[35] as

$$\rho_x(\mathbf{r},\mathbf{r}') = [\gamma(\mathbf{r},\mathbf{r}')]^2/2\rho(\mathbf{r}), \tag{61}$$

where

$$\gamma(\mathbf{r},\mathbf{r}') = \Sigma_i \psi_i^*(\mathbf{r})\psi_i(\mathbf{r}') \tag{62}$$

is the single-particle density matrix with $\gamma(\mathbf{r},\mathbf{r}) = \rho(\mathbf{r})$. In addition to satisfying the charge conservation sum rule

$$\int \rho_x(\mathbf{r},\mathbf{r}')\mathrm{d}\mathbf{r}' = 1, \tag{63}$$

the Fermi hole also satisfies the conditions

$$\rho_x(\mathbf{r},\mathbf{r}) = \rho(\mathbf{r})/2, \tag{64}$$

and

$$\rho_x(\mathbf{r},\mathbf{r}') \geq 0. \tag{65}$$

Thus, since the exact Fermi hole is known, it is possible to analyze both the local density and gradient expansion approximations for the exchange energy in terms of the corresponding hole charge distributions. Such an analysis would also lead to a better understanding of the limitations of these approximations for the correlation and exchange-correlation energies.

In the local density approximation (LDA), where the homogeneous electron gas pair-correlation function at the local value of the density is employed, the Fermi hole is given as[36]

$$\rho_x^{LDA}(\mathbf{r},\mathbf{r}') = (1/2)\rho(\mathbf{r})j^2(x),$$
$$j(x) = 3j_1(x)/x$$
$$x = k_F R \ , \ \mathbf{R} = \mathbf{r}' - \mathbf{r}, \tag{66}$$

where $j_1(x)$ is the first-order spherical Bessel function and $k_F(\mathbf{r}) = [3\pi^2\rho(\mathbf{r})]^{1/3}$ is the local value of the Fermi momentum. The LDA hole satisfies the constraints of Eqs. 63-65, but as a consequence of the assumption that each point be treated as a uniform electron gas system, the hole is always centered about the electron and spherically symmetric about it.

In Fig. 3 cross-sections through the exact and LDA Fermi holes in the electron-nucleus plane are plotted for the Neon atom for different electron positions using the analytical Hartree-Fock wavefunctions of Clementi and Roetti.[37] The positions considered are those[36] of the two maxima of the radial probability density at r = 0.104 a.u. and 0.651 a.u. (Figs. 3a and b), the intershell minimum at r = 0.307 a.u. (Fig. 3c), and a position in the classically forbidden region at r = 2.000 a.u. (Fig. 3d). The electron positions are indicated by arrows. (Note that the scale of Fig. 3a is an order or more greater in magnitude than those of the others).

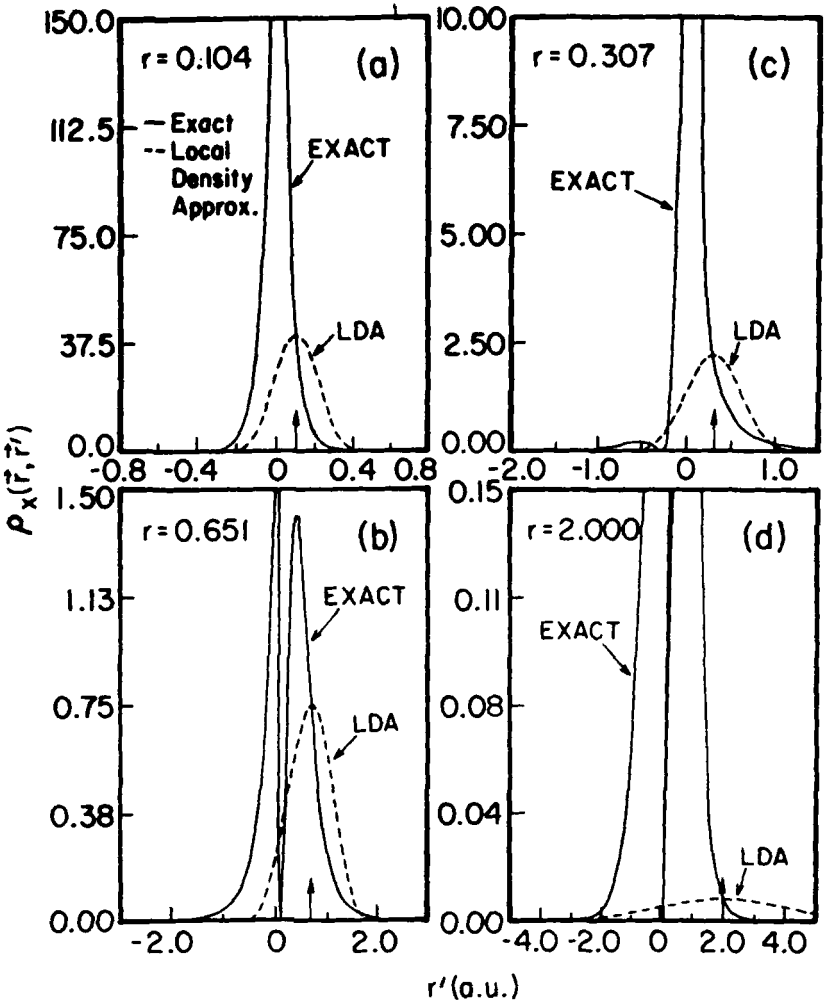


Fig. 3. Cross-sections through the exact and local density approximation (LDA) Fermi holes in the Neon atom for different positions r of the electron. (From Ref. 30).

TABLE I. Ground-state atomic energies in the Pauli-correlated approximation as obtained in the local density (LDA) and generalized gradient expansion (GGA) approximations, and within the Harbola-Sahni formalism and Hartree-Fock theory. The negative values of the energies in atomic units are quoted

ATOM	LDA	GGA[a]	HARBOLA-SAHNI[b]	HARTREE-FOCK[c]
Ar	524.517	526.935	526.804	526.818
Zn	1773.910	1778.183	1777.820	1777.848

a. See Ref. 48
b. See Ref. 55
c. See Ref. 8

TABLE II. Highest occupied eigenvalues of atoms in the Pauli-correlated approximation as obtained in the local density (LDA) and generalized gradient expansion (GGA) approximations, within the Harbola-Sahni formalism and Hartree-Fock theory as well as experimental ionization potentials. The negative values of the energies in Rydbergs are quoted.

ATOM	LDA	GGA[a]	HARTREE-FOCK[b]	HARBOLA-SAHNI[c]	EXPERIMENT[6]
Ar	0.668	0.686	1.182	1.178	1.158
Zn	0.371	0.390	0.585	0.646	0.690

a. See Ref. 48
b. See Ref. 8
c. See Ref. 55
d. See Ref. 38

In the interior and in particular the shell regions of the atom (Fig. 3a and b), the LDA hole is a fair approximation to the exact one. Since the principal contributions[36] to the exchange energy arise from the interior of the atom, the LDA should lead to accurate total energies for atoms. This is the case as indicated by the self-consistent results (see Table I) for Argon and Zinc for which the errors in comparison to Hartree-Fock theory[8] are 0.44% and 0.22% respectively. On the other hand, in the regions exterior to the atom and the asymptotic region, the LDA hole is poor (Fig. 3d) since it is centered about the electron and not about the nucleus as should be the case. Thus, the LDA potential in these regions is in error; it decays exponentially. Since the highest occupied eigenvalue depends principally on the structure of the effective potential in the region exterior to the atom, the LDA maximum eigenvalue will be in significant error. This is borne out by the results (see Table II) for Argon and Zinc on comparison with those of Hartree-Fock theory[8] and experimental[38] ionization potentials. The errors with respect to Hartree-Fock theory are 43.5% and 36.6% for Ar and Zn respectively.

The structure of the Fermi hole for the inhomogeneous electron gas at metallic surfaces is significantly different[39-41] from that in atoms. Whereas in atoms the hole is localized about the nucleus, it is delocalized and spreads throughout the crystal for electron positions near and outside the metal surface. It is from the region about and outside the surface that the principal contributions to the surface exchange energy arise.[36]. Since in the LDA the hole is centered and symmetric about the electron, over 50% of the hole in this approximation lies outside the crystal for these electron positions.[36] (For explicit comparisons of the exact and LDA holes we refer the reader to Ref. 36). This explains why the surface exchange energy in the LDA is in such significant

error. In Fig. 4 we show the variation of the exact[42] and LDA[42] surface exchange energies as a function of the density profile parameter y_F of the linear potential model.[43] Increasing values of y_F correspond to more slowly varying densities; $y_F = 0$ represents the infinite barrier model. Over the metallic range[44,45] of densities ($0.5<y_F<4.5$), the LDA is in error[42] by 46%-9%.

The Fermi hole in the gradient expansion approximation (GEA) is obtained[46] from an expansion[47] for the single-particle density matrix (and consequently the density) to second order in $v_{KS}(\mathbf{r})$, where $v_{KS}(\mathbf{r})$ is the local effective potential in which the electrons move. The expression for the GEA Fermi hole to $O(\nabla^2)$ is[46]

$$\rho_x^{GEA}(\mathbf{r},\mathbf{r}') = 1/2\ \rho(\mathbf{r})y(x),$$

$$y(x) = j^2(x) + [j(x)/4\pi^2][\mathbf{i}_R\cdot(\nabla k_F^2)\sin x + (\nabla k_F^2)^2\{L(x) - x\sin x\}/24k_F^3 + (\mathbf{i}_R\cdot\nabla k_F^2)^2\{x^2\cos x + 3\sin^2 x/j(x)\}/8k_F^3 - \nabla^2 k_F^2 L(x)/6k_F + (\mathbf{i}_R\cdot\nabla)^2 k_F^2 x\sin x)/3k_F]$$

$$L(x) = x^2 j(x)/3$$

$$\mathbf{i}_R = \mathbf{R}/R. \qquad (67)$$

Although the GEA hole satisfies the condition of Eq. 64, it satisfies neither the charge neutrality constraint (Eq. 63) nor the constraint that it be positive (Eq. 65). An improved hole can be obtained by imposing these constraints through appropriate configuration space cutoffs. Thus, the generalized gradient expansion (GGA) truncated hole may be written as[46]

$$\rho_x^{GGA}(\mathbf{r},\mathbf{r}') = 1/2\ \rho(\mathbf{r})y\Theta(y)\Theta[R_c(\mathbf{r}) - R], \qquad (68)$$

where the first step function ensures the hole is always positive and the second that charge neutrality is satisfied.

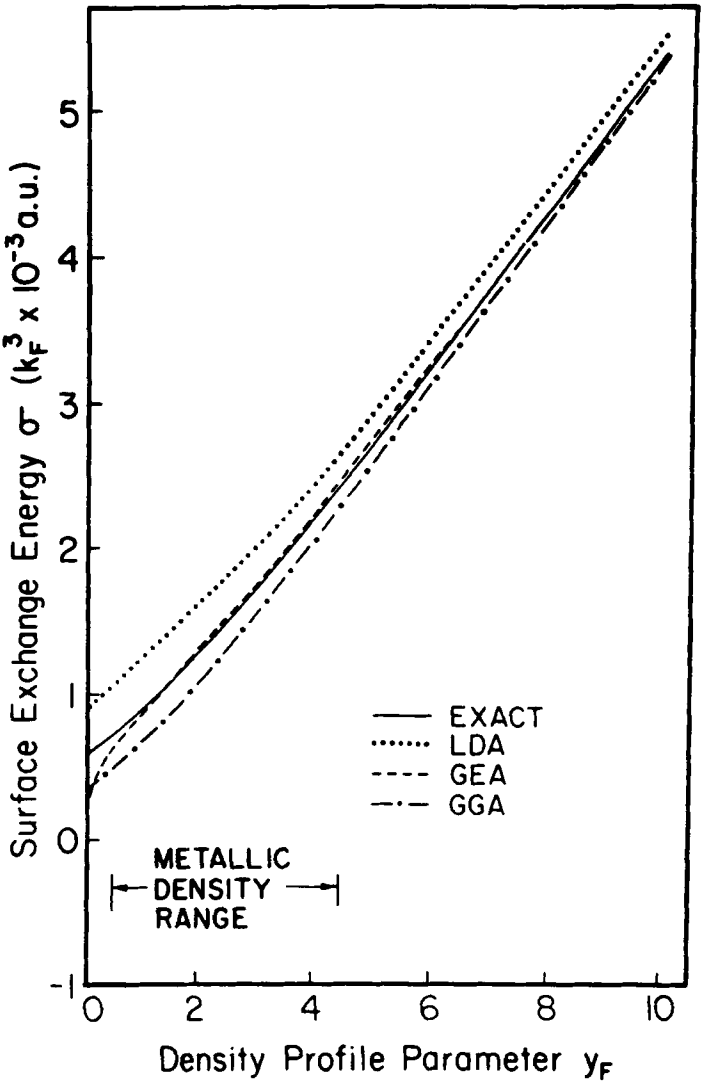


Fig. 4. Variation of the exact surface exchange energy as a function of the density profile parameter y_F of the linear potential model. Also plotted are the results in the local density (LDA), gradient expansion (GEA) and generalized gradient (GGA) approximations. The range of metallic density profiles is also indicated. (From Ref. 30).

In Fig. 5 cross-sections through the GGA hole for the Ne atom are plotted[30] for the same electron positions as in Fig. 3. Note the significant improvement of the GGA hole over the LDA hole in the interior of the atom (Fig. 5a,b,c). This improvement is reflected in the results[48] for the total energy (see Table I) where for Ar and Zn the GGA error is now only -0.02%. (Note that the GGA results lie below those of Hartree-Fock theory, whereas the energy of exchange-only density-functional theories should lie above due to the constraint[34] of locality imposed on the effective potential). Outside the atom Fig. 5d, the GGA hole begins to differ substantially from the exact result and in the asymptotic region reduces to the LDA hole. Thus, we do not expect much improvement over the LDA for the highest occupied eigenvalues. The improvement[48] (see Table II) is only 2% for Ar and 4% for Zn.

The GEA[14,49] and GGA[48] (with Sham's coefficient) surface exchange energies are also plotted in Fig. 4. The GGA results are an improvement over the LDA but underestimate the exact results; at $y_F = 0.5$ it is an underestimate by 33%; at $y_F = 4$ it reduces the 12% LDA error to one of 6.3%. The lack of as dramatic an improvement as in the case of atoms can be explained by the fact that in the regions about and outside the metal surface, the GGA hole becomes more like the LDA hole. Thus, instead of the hole spreading through the crystal, it is still localized about the electron. The GEA surface exchange energies, on the other hand are essentially exact for all densities for which $y_F \geq 1$. These results, in all probability, are fortuitous. The use of a coefficient which is 10/7ths that of Sham improves the results still further for these densities.

It is clear from the above analysis[30,36] that in the construction of an approximate hole charge distribution, the satisfaction of various sum rules is a necessary but not

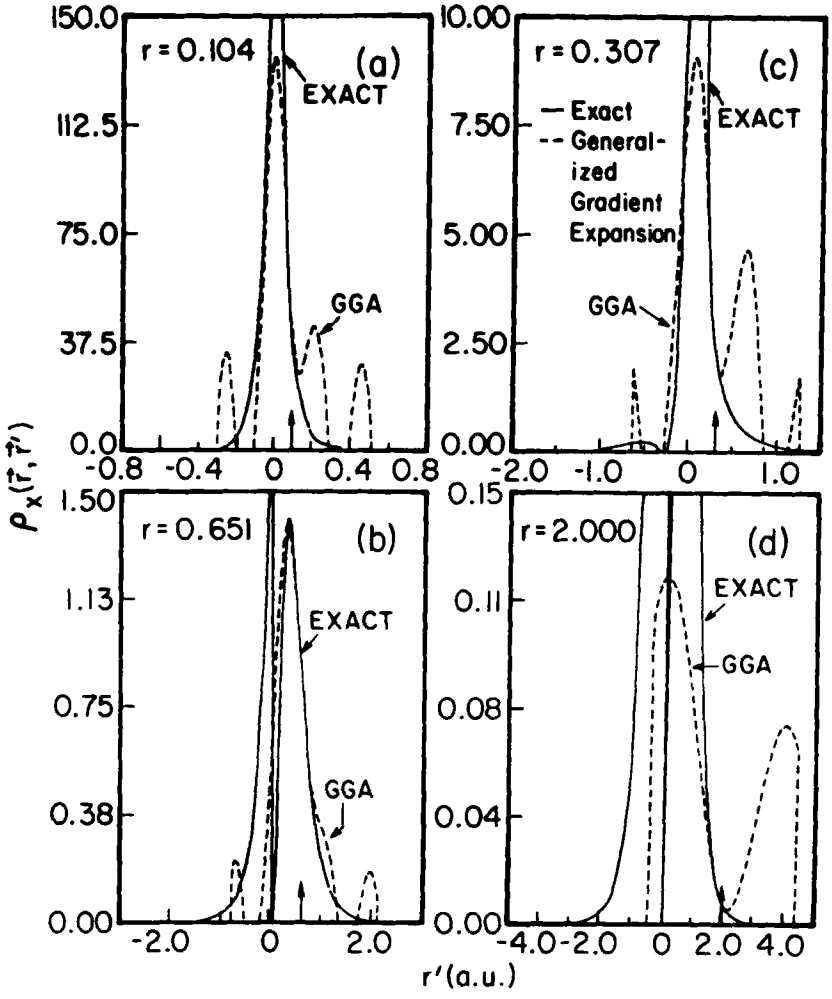


Fig. 5. Cross-sections through the exact and generalized gradient approximation (GGA) Fermi holes in the Neon atom for different positions r of the electron. (From Ref. 30).

sufficient requirement. What is required in addition is that approximate schemes replicate at least qualitatively the correct behavior of the hole distribution. For example, in the construction of an exchange energy functional, the corresponding Fermi hole must be localized for atoms and molecules, and delocalized for the non-uniform electron gas at metal surfaces. This will lead to good total energies and highest occupied eigenvalues.

As shown above, the construction of an accurate exchange-correlation energy functional by no means guarantees that its functional derivative will be too. The accuracy of the exchange-correlation potential is intrinsic to obtaining both ground-state energies as well as highest occupied eigenvalues correctly, but the latter property is particularly sensitive to the structure of this potential. In the classically forbidden region as the density becomes more slowly varying, any GEA (or GGA), whether it be for exchange-only or exchange-correlation, approaches the LDA. In the LDA, the Fermi-Coulomb hole is localized about the electron for all electron positions. Thus, no expansion in the gradients of the density to $O(\nabla^2)$ can lead to the correct asymptotic structure of the potential and thus to accurate highest occupied eigenvalues. They can, however, lead to good total energies.

Based on the fact that all the many-body effects are incorporated in the Fermi-Coulomb hole, Harbola and Sahni[50] have recently proposed a physical interpretation for the exchange-correlation potential whereby it may be obtained directly from the hole charge distribution, thus obviating the need to determine functional derivatives. Since the Fermi-Coulomb hole is dynamic as a function of electron position, the exchange-correlation potential is the work $W_{xc}(\mathbf{r})$ required to bring an electron from infinity to its final position against the force field of this charge distribution. Thus

$$W_{xc}(\mathbf{r}) = -\int_{\infty}^{\mathbf{r}} \boldsymbol{\varepsilon}_{xc} \cdot d\mathbf{l}, \tag{69}$$

where $\boldsymbol{\varepsilon}_{xc}(\mathbf{r})$ is the electric field due to the Fermi-Coulomb hole which according to Coulomb's law is

$$\boldsymbol{\varepsilon}_{xc}(\mathbf{r}) = \int d\mathbf{r}' \; \rho_{xc}(\mathbf{r},\mathbf{r}')(\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3. \tag{70}$$

It is implicit in this interpretation that the potential $W_{xc}(\mathbf{r})$ is path independent or equivalently that the curl of the electric field $\boldsymbol{\varepsilon}_{xc}(\mathbf{r})$ vanishes. For spherically symmetric systems,[50] the electric field is in the radial direction and a function only of $|\mathbf{r}|$, the distance from the origin. For jellium metal surfaces,[50,51] the electric field is in the direction perpendicular to the surface and a function of the perpendicular co-ordinate variable only. Thus, in these cases the curl vanishes. For systems in which the curl of the electric field is not zero, the potential may be constructed from the longitudinal component of the electric field, thus ensuring the path independence of the potential.

With the above interpretation the exchange potential $W_x(\mathbf{r})$ can be determined exactly since the Fermi hole is known explicitly in terms of the orbitals. Another consequence of this interpretation is that as the Fermi-Coulomb and Fermi hole charge distributions both satisfy the charge conservation sum rule, the total Coulomb hole charge is zero.[52] Thus, asymptotically the Coulomb hole does not contribute to the potential. The asymptotic structure of the potential for the fully-correlated system is thus that of W_x alone and therefore known exactly. For atoms, the asymptotic structure of the potential W_x has been shown[50] to be $-1/r$, and that at metallic surfaces[51] to be the image potential $-1/4x$. Furthermore, at metal surfaces, the potential W_x (in units of $3k_F/2\pi$) relative to the vacuum level goes[50] to the Kohn-Sham homogeneous

electron gas value of -2/3 as it must. The local exchange potential $W_x\{\psi_i\}$ when substituted for the functional derivative $v_x\{\rho(\mathbf{r})\} = \delta E_x[\rho]/\delta\rho$ also satisfies[50] exactly the virial theorem based sum rule[53]

$$E_x[\psi_i] = 1/2 \sum_{ij(\uparrow\uparrow \text{ spins})} \int\int d\mathbf{r}d\mathbf{r}' \psi_i^*(\mathbf{r})\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')\psi_j(\mathbf{r})/|\mathbf{r}-\mathbf{r}'| = -\int d\mathbf{r}\rho(\mathbf{r})\mathbf{r}\cdot\nabla v_x\{\rho(\mathbf{r})\}. \quad (71)$$

Furthermore, the potential W_x scales as $W_x[\rho_\lambda] = \lambda W_x[\rho]$, where $\psi_{i,\lambda}(\mathbf{r}) = \lambda^{3/2}\psi_i(\lambda\mathbf{r})$ and $\rho_\lambda(\mathbf{r}) = \lambda^3\rho(\lambda\mathbf{r})$, so that it also satisfies the scaling sum rule[54]

$$(\partial/\partial\lambda)\int f_\lambda(\mathbf{r})v_x\{\rho_\lambda(\mathbf{r})\}d\mathbf{r}\Big|_{\lambda=1} = \int f(\mathbf{r})v_x\{\rho(\mathbf{r})\}d\mathbf{r}, \quad (72)$$

where the function f is such that $\int f(\mathbf{r})d\mathbf{r} = 0$ but with the additional constraint that $\rho+f\geq 0$. The self consistently determined[55] ground-state energies and highest occupied eigenvalues with the W_x potential for Ar and Zn are also quoted in Tables I and II. The total energies lie above those of Hartree-Fock theory as they must[34], and differ from it by less than 50ppm. Since the highest occupied eigenvalue of density-functional theory is minus the ionization potential, and as the asymptotic structure of the exchange-correlation potential is that of W_x, one expects that the highest occupied eigenvalues in exchange-only theory should be close to those of the fully-correlated system. Thus, comparisons with experiment are meaningful within this formalism even in the exchange-only case. Observe that the W_x-formalism highest occupied eigenenergies closely approximate the experimental values, which indicates that the potential is accurate not only asymptotically, but also in regions closer to the atom. Thus,

we see that from the Fermi-Coulomb hole charge distribution of an electron, both the exchange-correlation energy as well as the highest occupied eigenvalue of interacting non-uniform electrons can be obtained accurately.

The question of whether the potential W_x is the true Kohn-Sham potential cannot be answered at present. The orbitals of the Slater determinant employed in the current density-functional theory definition[14,34] of the exchange energy are those that minimize the expectation value of the Hamiltonian and which simultaneously are the ground-state eigenfunctions of a non-interacting local potential Hamiltonian. This optimized effective potential must[53] in principle satisfy the virial theorem based sum rule of Eq. (71). The fact[50] that neither the Talman et al[9] nor the HOvD[7] potentials satisfy this sum rule, together with the fact that they lead to the same total ground-state energies, implies that neither is the Kohn-Sham potential. They do, however, lead to energies that are lower than those of the W_x formalism. On the other hand, the potential W_x is the only potential known which satisfies[50] analytically and numerically the virial theorem sum rule in orbital form. It also satisfies all scaling requirements. (We note that various approximate energy functionals such as that of the LDA for example, also satisfy the virial theorem sum rule but in an approximate manner. By this is meant that the approximate energy functional appears on the left side of Eq. (71), and its functional derivative on the right). Thus, even if the potential W_x is not the exact functional derivative of the exchange energy functional, the results of application to both few-electron atomic[50,55] and many-electron metallic surface[50,51] non-uniform systems clearly demonstrates the high accuracy of this potential. Finally, we conclude by noting that since in the LDA the Fermi-Coulomb hole is symmetrical about the electron, the LDA potential obtained

via the above interpretation vanishes throughout space.[50] The GGA potential, however, is finite since the corresponding hole is not centered and symmetric about the electron for all electron positions.[30]

REFERENCES

1. P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
2. W. Kohn and L.J. Sham, Phys. Rev. 140, A1133 (1965).
3. D. Pines and P. Nozieres, The Theory of Quantum Liquids, Vol. 1 (Addison Wesley, California (1989)).
4. J.P. Perdew, R.G. Parr, M. Levy and J.L. Balduz, Phys. Rev. Lett. 49, 1691 (1982).
5. M. Levy, J.P. Perdew and V. Sahni, Phys. Rev. A30, 2745 (1984).
6. S.-K. Ma and K.A. Brueckner, Phys. Rev. 165, 18 (1968).
7. F. Herman, J.P. Van Dyke and I.B. Ortenburger, Phys. Rev. Lett. 22, 807 (1969); F. Herman, I.B. Ortenburger and J.P. Van Dyke, Int. J. Quant. Chem. IIIS, 827 (1970).
8. C.F. Fischer, The Hartree-Fock Method for Atoms (Wiley, New York, 1977).
9. J.D. Talman and W.F. Shadwick, Phys. Rev. A14, 36 (1976); K. Aashamar, T.M. Luke and J.D. Talman, At. Data Nucl. Data Tables 22, 443 (1978); R.T. Sharp and G.K. Horton, Phys. Rev. 90, 317 (1953).
10. L.J. Sham, in Computational Methods in Band Theory, edited by P. Marcus, J.F. Janak and A.R. Williams (Plenum, New York, 1971).
11. L. Kleinman, Phys. Rev. B10, 2221 (1974); 12, 3512 (E) (1975).
12. D.J.W. Geldart, M. Rasolt and C.O. Ambladh, Solid State Commun. 16, 243 (1975).
13. A.K. Rajagopal and S. Ray, Phys. Rev. B12 3129 (1975).
14. V. Sahni, J. Gruenebaum and J.P. Perdew, Phys. Rev. B26, 4371 (1982).
15. A.R.E. Mohammed and V. Sahni, Phys. Rev. B29, 3687 (1984).
16. J.E. Robinson, F. Bassani, R.S. Knox and J.R. Schrieffer, Phys. Rev. Lett. 9, 215 (1962).
17. D.C. Langreth and J.P. Perdew, Phys. Rev. B21, 5469 (1980).
18. V. Peuckert, J. Phys. C9, 4173 (1976).
19. D.J.W. Geldart and M. Rasolt, Phys. Rev. B13, 1477 (1976).

20. We have put the LP result into Hartree units and multiplied by a factor of 2 to make it consistent with our Eq. (10) which differs by a factor of 1/2 from the corresponding equation of LP (Ref. 17).
21. L. Kleinman, Phys. Rev. B30, 2223 (1984).
22. P.R. Antoniewicz and L. Kleinman, Phys. Rev. B31, 6779 (1985).
23. D.C. Langreth and S.H. Vosko, Phys. Rev. Lett. 59, 497 (1987).
24. J.P. Perdew, private communication and J.P. Perdew and Y. Wang, in Mathematics Applied to Science, edited by J. Goldstein, S. Rosencrans and G. Sod (Academic, New York, 1987).
25. L. Kleinman and S. Lee, Phys. Rev. B37, 4634 (1988).
26. J.A. Chevary and S.H. Vosko, Bull. Am. Phys. Soc. 33, 238 (1988).
27. D.J.W. Geldart and R. Taylor, Can. J. Phys. 48, 155 (1970).
28. L. Kleinman and T. Tamura, Phys. Rev. B40, xxx (1989).
29. J.P. Perdew, private communication.
30. M.K. Harbola and V. Sahni, (manuscript in preparation).
31. J. Harris and R.O. Jones, J. Phys. F4, 1170 (1974).
32. O. Gunnarsson and B.I. Lundqvist, Phys. Rev. B13, 4274 (1976).
33. D.C. Langreth and J.P. Perdew, Solid State Commun. 17, 1425 (1975); Phys. Rev. B15, 2884 (1977).
34. V. Sahni and M. Levy, Phys. Rev. B33, 3869 (1986).
35. J.C. Slater, Phys. Rev. 81, 385 (1951).
36. V. Sahni, K.-P. Bohnen and M.K. Harbola, Phys. Rev. A37, 1895 (1988).
37. E. Clementi and C. Roetti, At. Data Nucl. Data Tables 14, 177 (1974).
38. C.E. Moore, Ionization Potentials and Ionization Limits Derived from the Analysis of Optical Spectra, Nat. Stand. Ref. Data Ser., Nat. Bur. Stand. (U.S.), 34, (1970).
39. V. Sahni and K.-P. Bohnen, Phys. Rev. B29, 1045 (1984).
40. V. Sahni and K.-P. Bohnen, Phys. Rev. B31, 7651 (1985).
41. M.K. Harbola and V. Sahni, Phys. Rev. B37, 745 (1988).
42. C.Q. Ma and V. Sahni, Phys. Rev. B20, 2291 (1979). More numerically refined values for the exact surface exchange energy are quoted in Ref. 14.
43. V. Sahni, J.B. Krieger and J. Gruenebaum, Phys. Rev. B15 1941 (1977); V. Sahni and J. Gruenebaum, Solid State Commun. 21, 463 (1977).
44. V. Sahni and J. Gruenebaum, Phys. Rev. B19, 1840 (1979).
45. V. Sahni, J.P. Perdew and J. Gruenebaum, Phys. Rev. B23, 6512 (1981).
46. J.P. Perdew, Phys. Rev. Lett. 55, 1665 (1985); J.P. Perdew and W. Yue, Phys. Rev. B33, 8800 (1986).

47. E.K.U. Gross and R.M. Dreizler, Z. Phys. A302, 103 (1981).
48. J.P. Perdew, M.K. Harbola and V. Sahni, in Condensed Matter Theories, vol. 3, 235 (Edited by J.S. Arponen, R.F. Bishop and M. Manninen, Plenum 1988).
49. V. Sahni and J. Gruenebaum, Phys. Rev. B25, 6275 (1982).
50. M.K. Harbola and V. Sahni, Phys. Rev. Lett. 62, 489 (1989).
51. M.K. Harbola and V. Sahni, Phys. Rev. B39, 10437 (1989); V. Sahni, Surf. Sci. 213, 226 (1989).
52. Note that this definition of the Coulomb hole differs from that employed in the initial section of this paper on the coefficients of the first gradient correction term.
53. M. Levy and J.P. Perdew, Phys. Rev. B32, 2010 (1985).
54. M. Levy, private communication.
55. Y. Li, M.K. Harbola, J.B. Krieger and V. Sahni, submitted to Phys. Rev. A.

MAGNETIC FIELDS AND DENSITY FUNCTIONAL THEORY

G. Vignale

Department of Physics, University of Missouri-Columbia
Columbia, Missouri 65211 U.S.A.

Mark Rasolt

Solid State Division, Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6032 U.S.A.

D.J.W. Geldart

Department of Physics, Dalhousie University
Halifax, Nova Scotia, Canada B3H3J5

1. INTRODUCTION

Since the seminal papers of Hohenberg and Kohn[1] and Kohn and Sham[2], density-functional theory (DFT) has developed into an important tool for the treatment of many-body problems in condensed matter physics. Its practical success in the calculation of the electronic structure has prompted much theoretical work aimed at extending the applicability of the theory to systems more general than those considered in the original papers[3,4].

Due to the importance of magnetic fields as probes of condensed matter, and, perhaps more crucially, due to the occurrence in nature of ferromagnets which spontaneously break spin rotational symmetry, efforts to bring magnetic fields within the range of action of DFT were made as early as 1973[5]. It was recognized that, in the presence of a magnetic field $\vec{B}$, the mapping between ground-state energy and ground-state wavefuntion is not one to one, i.e., the Hohenberg-Kohn theorem does not hold in this case. To recover a one to one mapping it is necessary to supplement the number density distribution by spin- and current-densities which play the same role with respect to an external field $\vec{B}$, and vector potential $\vec{A}$, as

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

the density does with respect to the ordinary scalar potential. In practice, however, the orbital currents were ignored - perhaps due to the conceptual difficulties described below - and the interest focussed on the effects of spin-polarization which were easily incorporated in a self-consistent Kohn-Sham scheme, now known as "spin-density functional theory" (SDFT)[3,6,7].

Obviously, the above approach is not justified when the magnetic field is sufficiently strong. For instance, a large $\vec{B}$ field can completely polarize the electrons in the lowest Landau level: in this regime, the current-dependent part of the correlation energy gives the decisive contribution to the ground-state energy and actually determines the structure of the ground-state itself[8]. Generally speaking, inclusion of currents in density functional theory is required whenever a large magnetic field coexists with a strongly inhomogeneous electronic structure. Concrete examples of this situation are: (i) broken-symmetry states (charge-density waves, spin-density waves, Wigner crystals) in three-dimensional semimetals and doped semiconductors in strong magnetic fields[9] (ii) broken-symmetry states in the fractional quantum-Hall effect regime of a two-dimensional electron gas[10,11] (iii) ground-state energy calculation on finite size systems (such as atoms, molecules, small metal particles) in strong magnetic fields. However, until recently the application of the density-functional method to such physical situations was at best tentative with regard to its basic principles[11].

Although simple in appearance, constructing an acceptable non-relativistic current-density-functional theory (CDFT) is a remarkably delicate task. One can understand the basic difficulty by considering the non-relativistic expression for the orbital current density

$$\vec{j}(\vec{r}) = \vec{j}_p(\vec{r}) + (e/mc)n(\vec{r})\vec{A}(\vec{r}) \tag{1.1}$$

The first term $\vec{j}_p(\vec{r})$ is referred to as "paramagnetic current density," and measures the density of canonical momentum in the ground-state wavefunction. The second term is referred to as "diamagnetic current density" and depends explicitly on the number density $n(\vec{r})$ and the vector potential $\vec{A}(\vec{r})$. In the application of the variational principle the vector potential must be kept constant. It follows that the variation of $\vec{j}$ (at constant n) is effectively a variation with respect to $\vec{j}_p$: therefore $\vec{j}_p$ not $\vec{j}$, must be used as basic variable. All this makes perfect sense physically since $\vec{j}$ contains the relevant information about the electron wavefunction (for instance, in the homogeneous electron gas, in a uniform $\vec{B}$ field, $\vec{j}$ vanishes everywhere, while at the same time, $\vec{j}_p$ is huge!). Unfortunately $\vec{j}_p$ has the serious drawback of not being a gauge-invariant quantity. Thus, the challenge is to formulate a gauge-invariant theory in terms of non gauge-invariant variables.

Recently, Vignale and Rasolt (VR)[12,13] succeeded in constructing a non-relativistic $CDFT$ which, using $\vec{j}_p$ as one of the basic variables, satisfies the physical requirement of gauge-invariance and the continuity equation for the current. The cornerstone of the new theory is the recognition of a fundamental symmetry of the exchange-correlation energy functional E_{xc}, in virtue of which the latter is actually a functional of the gauge-invariant combination

$$\vec{v}(\vec{r}) = \vec{\nabla} \times [(\vec{j}_p(\vec{r}))/n(\vec{r})] \tag{1.2}$$

The physical meaning of the new variable $\vec{v}(\vec{r})$ will be discussed in the following sections.

Most of this chapter is devoted to a detailed presentation of the VR formalism. In Section 2 we justify the choice the basic variables of this theory, and prove the analogues of the Hohenberg-Kohn theorem and the variational principle. In Section 3 we present the derivation of the self-consistent equations, and discuss their unusual structure. In Section 4 the intimately related issues of gauge-invariance, current conservation, and crystal translational symmetry are examined, showing that the VR theory is well-behaved on all three counts. In section 5 we introduce the general principle of the local-density approximation, and provide a new derivation of its weak-field limit. Section 6 reviews the status of the unification of spin- and current-density functional theory. Finally, Section 7 presents a review of open questions and recent developments in this field.

2. BASIC VARIABLES AND THEOREMS

2.1 Hohenberg-Kohn Theorem

We consider the non-relativistic Hamiltonian for N electrons (e is the absolute value of the charge)

$$H = H_0 + \int n^{op}(\vec{r})[V(\vec{r}) + \frac{e^2}{2mc^2}A^2(\vec{r})]dr + \frac{e}{c}\int \vec{j}_p^{op}(\vec{r}) \cdot \vec{A}(\vec{r})dr \quad . \tag{2.1}$$

H_0 is the Hamiltonian of the homogeneous electron gas,

$$H_0 = -\sum_{i=1}^{N} \frac{\hbar^2 \nabla_i^2}{2m} + \frac{1}{2}\sum_{\substack{i,j \\ i \neq j}} \frac{e^2}{\mid \vec{r}_i - \vec{r}_j \mid} \quad . \tag{2.2}$$

The remaining terms describe interactions with external scalar and vector potentials $V(\vec{r})$ and $\vec{A}(\vec{r})$. In order to emphasize the novel feature of the orbital current we ignore, for the time being, the spin. The inclusion of

spin will be reviewed in Section 6. The number density and paramagnetic current density operators in Eq. (2.1) are defined as follows:

$$n^{op}(\vec{r}) = \psi^+(\vec{r})\psi(\vec{r})$$
$$\vec{j}_p^{op}(\vec{r}) = -\frac{i\hbar}{2m}\{\psi^+(\vec{r})\vec{\nabla}\psi(\vec{r}) - [\vec{\nabla}\psi^+(\vec{r})]\psi(\vec{r})\} \quad . \tag{2.3}$$

The corresponding ground-state expectation values are denoted by $n(\vec{r})$ and $\vec{j}_p(\vec{r})$. The physical current density is given by Eq. (1.1). The constancy of the total number of electron and the continuity equation for the current $[\vec{\nabla}\cdot\vec{j}(\vec{r}) = 0]$ impose the constraints

$$\int n(\vec{r})d\vec{r} = N$$
$$\vec{\nabla}\cdot\vec{j}_p(\vec{r}) + (e/mc)\vec{\nabla}\cdot[n(\vec{r})\vec{A}(\vec{r})] = 0 \quad . \tag{2.4}$$

The generalized Hohenberg-Kohn theorem states that $V(\vec{r})$ and $A(\vec{r})$, and the non-degenerate ground-state wavefunction $\psi(\vec{r}_1 \cdots \vec{r}_N)$ are uniquely determined (apart from an additive constant in V) by the distributions $n(\vec{r})$ and $\vec{j}_p(\vec{r})$. The proof is by *reductio ad adsurdum*. Suppose that there are two sets of fields $V(\vec{r}), \vec{A}(\vec{r})$ and $V'(\vec{r}), \vec{A}'(\vec{r})$ giving the same ground state distributions $n(\vec{r})$ and $\vec{j}_p(\vec{r})$[14]. Let ψ and ψ' be the two different ground states corresponding to the two sets of fields. Let H and H' be the two corresponding Hamiltonians and E and E' the two ground-state energies. Then, from the variational principle for the ground-state of H, we obtain the inequality

$$E = \langle\psi \mid H \mid \psi\rangle < \langle\psi' \mid H \mid \psi'\rangle = E' + \int n(\vec{r})\Big[V(\vec{r}) + \frac{e^2}{2mc^2}A^2(\vec{r})$$
$$-V'(\vec{r}) - \frac{e^2}{2mc^2}A'^2(\vec{r})\Big]d\vec{r} + \frac{e}{c}\int \vec{j}_p(\vec{r})[\vec{A}(\vec{r}) - \vec{A}'(\vec{r})]d\vec{r}.$$

Another inequality is obtained by interchanging the primed and unprimed variables, and summing the two inequalities we get the contradiction

$$E + E' < E + E' \tag{2.5}$$

which proves the theorem. Note that this theorem would have not been true, had one attempted to work with the physical current density $\vec{j}$ as basic variable. In that case one would still have the freedom of operating a gauge transformation on the vector potential, and thus multiplying the ground-state wavefunction by a coordinate-dependent phase factor, without

changing the physical current density. Therefore, the ground-state wave function is *not* a unique functional of $\vec{j}$. Working with the paramagnetic current density eliminates this ambiguity.

2.2. Variational principle

The variational principle of $CDFT$ is formulated as follows. Let

$$F[n,\vec{j}_p] \equiv \langle \psi[n,\vec{j}_p] \mid H_o \mid \psi[n,\vec{j}_p] \rangle, \tag{2.6}$$

where $\psi[n,\vec{j}_p]$ is the ground-state wavefunction corresponding to n and $\vec{j}_p$, be the "internal energy" functional. Then the functional

$$E_{VA}[n',\vec{j}'_p] = F[n',\vec{j}'_p] + \int n'(\vec{r})\left[V(\vec{r}) + \frac{e^2}{2mc^2}A^2(\vec{r})\right]d\vec{r} + \frac{e}{c}\int \vec{j}'_p(\vec{r})\vec{A}(\vec{r})d\vec{r} \tag{2.7}$$

where V and $\vec{A}$ are given external fields, has a minimum when n' and j'_p take the correct values corresponding to the potentials V and $\vec{A}$. This minimum value is the correct ground-state energy of the system in the presence of V and $\vec{A}$. The proof of this theorem follows from the variational principle for the ground-state of H, since

$$\begin{aligned} E_{VA}[n',\vec{j}'_p] &\equiv \langle \psi[n',j'_p] \mid H \mid \psi[n',j'_p] \rangle \geq \langle \psi[n,\vec{j}_p] \mid H \mid \psi[n,\vec{j}_p] \rangle \\ &\equiv E_{VA}[n,\vec{J}_p] \end{aligned} \tag{2.8}$$

which proves the theorem.

3. SELF-CONSISTENT EQUATIONS

In this section we show that the minimization of $E_{VA}[n',\vec{j}'_p]$ is formally equivalent to the self-consistent solution of a one-electron Schroedinger-type equation.

We first decompose the internal energy functional as follows:

$$F[n',\vec{j}'_p] = T_s[n',\vec{j}'_p] + \frac{e^2}{2}\int\int \frac{n'(\vec{r})n'(\vec{r}')}{\mid \vec{r}-\vec{r}' \mid}d\vec{r}d\vec{r}' + E_{xc}[n',\vec{j}'_p] \tag{3.1}$$

The first term T_s is the non-interacting version of the F functional:

$$T_s[n',\vec{j}'_p] \equiv \langle \psi^0[n',\vec{j}'_p] \mid -\sum_{i=1}^{N}\frac{\hbar^2\nabla_i^2}{2m} \mid \psi^0[n',\vec{j}'_p] \rangle \tag{3.2}$$

where $\psi^0[n', \vec{j}'_p]$ is the ground-state wavefunction of N non-interacting electrons with the given density and current distributions. The second term is the classical electrostatic energy. The third term E_{xc} is, by definition, the exchange-correlation energy functional.

By construction $\psi^0[n', \vec{j}'_p]$ is a Slater determinant of N one-electron orbitals, which are the N lowest lying solutions of a Schroedinger equation with some appropriate external fields $V_{eff}[n', \vec{j}'_p]$ and $A_{eff}[n', \vec{j}'_p]$[15] :

$$\left[\frac{1}{2m}\left(-i\hbar\vec{\nabla} + \frac{e}{c}\vec{A}_{eff}(\vec{r})\right)^2 + V_{eff}(\vec{r})\right]\psi'_i(\vec{r}) = \epsilon'_i\psi'_i(\vec{r}) \quad . \tag{3.3}$$

The densities n' and $\vec{j}'_p$ are related to the one-electron orbitals by the equations

$$\begin{aligned} n'(\vec{r}) &= \sum_{i=1}^{N} \mid \psi'_i(\vec{r}) \mid^2 \\ \vec{j}'_p(\vec{r}) &= \sum_{i=1}^{N}\left(-\frac{i\hbar}{2m}\right)\left\{\psi'^*_i(\vec{r})\vec{\nabla}\psi'_i(\vec{r}) - [\vec{\nabla}\psi'^*_i(\vec{r})]\psi'_i(\vec{r})\right\} \end{aligned} \tag{3.4}$$

The functional T_s can now be expressed in terms of the one electron eigenvalues ϵ'_i. Multiplying Eq. (3.3) to the left by ψ'^*_i, integrating over $\vec{r}$, and summing over i one finds

$$\begin{aligned} T_s[n', \vec{j}'_p] = \sum_{i=1}^{N}\epsilon'_i - \int n'(\vec{r})\left[V_{eff}(\vec{r}) + \frac{e^2}{2mc^2}A^2_{eff}(\vec{r})\right]d\vec{r} \\ - \frac{e}{c}\int \vec{j}'_p(\vec{r})\cdot\vec{A}_{eff}(\vec{r})d\vec{r} \end{aligned} \tag{3.5}$$

Combining Eqs. (2.8), (3.1) and (3.5) we construct the functional E_{VA} to be minimized. In calculating the functional derivatives of $E_{VA}[n', \vec{j}'_p]$ with respect to n' and $\vec{j}'_p$ one must keep in mind that V_{eff} and A_{eff} are functionals of these distributions. However, as a consequence of Eq. (3.3), the derivatives of ϵ'_i are exactly cancelled by those of V_{eff}. Thus, the variational principle leads to the equations

$$V_{eff}(\vec{r}) = V(\vec{r}) + V_H(\vec{r}) + V_{xc}(\vec{r}) + \frac{e^2}{2mc^2}[A^2(\vec{r}) - A^2_{eff}(\vec{r})] \tag{3.6}$$

and

$$A_{eff}(\vec{r}) = \vec{A}(\vec{r}) + \vec{A}_{xc}(\vec{r}) \tag{3.7}$$

where

$$V_H(\vec{r}) = e^2 \int \frac{n'(\vec{r}')}{\mid \vec{r} - \vec{r}' \mid} d\vec{r}' \quad ,$$

$$V_{xc}(\vec{r}) = [\delta E_{xc}[n', \vec{j}_p']/\delta n'(\vec{r})]_{\vec{j}_p'} \quad ,$$

$$\vec{A}_{xc}(\vec{r}) = [\delta E_{xc}[n', \vec{j}_p']/\delta \vec{j}_p'(\vec{r})]_{n'} \quad . \tag{3.8}$$

Equations (3.6) to (3.8) determine the fields V_{eff} and $\vec{A}_{eff}$ as functionals on n' and j_p'. The latter are in turn determined by V_{eff} and $\vec{A}_{eff}$ via equations (3.3) and (3.4). Thus, we have obtained a self-consistent set of one-electron equations, from which the ground-state properties can be calculated.

At first sight, equation (3.3) appears to have the conventional structure of a genuine one-electron Schroedinger problem in the presence of a vector potential $\vec{A}_{eff}$. But a closer inspection reveals a striking difference: the effective scalar potential is now also dependent on $\vec{A}$ and $\vec{A}_{eff}$. One can easily verify that the vector potential part of V_{eff} leads to an equation in which $\vec{A}_{eff}$ appears linearly: the quadratic term remains $(e^2/2mc^2)A^2$ as in the non-interacting case. The explicit form of this equation is

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 - \frac{i\hbar e}{mc}[\vec{A}_{eff}(\vec{r}) \cdot \vec{\nabla} + \vec{\nabla} \cdot \vec{A}_{eff}(\vec{r})] \right.$$
$$\left. + \frac{e^2}{2mc^2} A^2(\vec{r}) + V(\vec{r}) + V_H(\vec{r}) + V_{xc}(\vec{r}) \right\} \psi_i(\vec{r}) = \epsilon_i \psi_i(\vec{r}) \tag{3.9}$$

The ground-state energy can be obtained inserting the expressions for V_{eff} and $\vec{A}_{eff}$ given in Eq. (3.7) in T_s and E_{VA}. The result is

$$E_G = \sum_{i=1}^{N} \epsilon_i - \frac{1}{2} \int \int \frac{n(\vec{r}) n(\vec{r}')}{\mid \vec{r} - \vec{r}' \mid} d\vec{r} d\vec{r}' - \int n(\vec{r}) V_{xc}(\vec{r}) d\vec{r}$$
$$- \frac{e}{c} \int \vec{j}_p(\vec{r}) \cdot \vec{A}_{xc}(\vec{r}) dr + E_{xc}[n, \vec{j}_p] \quad . \tag{3.10}$$

We note, in passing, that the chemical potential μ, determined as the largest occupied eigenvalues of the Schroedinger-like equation, is exact (Koopman's theorem of $CDFT$).

4. GAUGE-INVARIANCE

The unusual structure of the self-consistent equations derived in the previous section poses a number of serious questions. Consider the gauge transformation

$$\vec{A}(\vec{r}) \rightarrow \vec{A}(\vec{r}) - \vec{\nabla}\Lambda(\vec{r}) \quad , \tag{4.1}$$

where $\Lambda(\vec{r})$ is an arbitrary function. Obviously, all the physical quantities, like $n(r), \vec{j}(r)$, and E_G must be invariant under this transformation. Does Eq. (3.9) satisfy this requirement? A second question arises in connection with the continuity equation constraint (Eq. 2.4). The solution of Eq. (3.3) appears - according to the usual properties of the Schroedinger equation - to satisfy a different form of the constraint whereby $\vec{A}$ is replaced by $\vec{A}_{eff}$. How can this be correct?

In order to answer the above questions it is essential to determine the behavior of the various functions under the transformation

$$\vec{j}_p(\vec{r}) \rightarrow \vec{j}_p(\vec{r}) + (e/mc)n(\vec{r})\vec{\nabla}\Lambda(\vec{r}) \quad . \tag{4.2}$$

We begin with the internal energy functional F defined in Eq. (2.6). The ground-state wavefunction transforms according to the rule

$$\psi[n', \vec{j}_p'] \rightarrow exp\left[i\frac{e}{\hbar c}\sum_{i=1}^{N}\Lambda(\vec{r}_i)\right]\psi[n', \vec{j}_p'] \tag{4.3}$$

(The validity of this equation can be directly verified by noting that Eq. (4.2) is just a gauge transformation for $\vec{j}_p$.) From this, the transformation of F is immediately derived:

$$F[n', \vec{j}_p'] \rightarrow F[n', \vec{j}_p'] + \frac{e}{c}\int \vec{j}_p'(\vec{r})\cdot\vec{\nabla}\Lambda(\vec{r})d\vec{r} + \frac{e^2}{2mc^2}\int n'(\vec{r}) \mid \vec{\nabla}\Lambda(\vec{r}) \mid^2 \quad . \tag{4.4}$$

The crucial fact about this equation is that the transformation depends only on $\vec{j}_p'$ and n', not on the details of the ground-state wavefunction. Therefore the same transformation must also apply to the non-interacting functional T_s. Taking the difference of F and T_s and using Eq. (3.1) leads to the important result

$$E_{xc}[n', \vec{j}_p' + \frac{e}{mc}n'\vec{\nabla}\Lambda] = E_{xc}[n', \vec{j}_p'] \tag{4.5}$$

for arbitrary Λ. A compact way of expressing the above invariance is to say that E_{xc} is actually a functional of the gauge-invariant combination $\vec{v}$ defined in Eq. (1.2)

$$E_{xc}[n, \vec{j}_p] = \overline{E}_{xc}[n, \vec{v}] \quad . \tag{4.6}$$

Equation (4.6) is at the heart of the VR formulation. Differentiating both sides of Eq. (4.6) with respect to n and $\vec{j}_p$ and using the definitions of Eqs. (3.8) one finds the following expression for the effective potentials:

$$V_{eff}(\vec{r}) = V(\vec{r}) + V_H(\vec{r}) + \frac{\delta\overline{E}_{xc}[n, \vec{v}]}{\delta n(\vec{r})}\bigg|_{\vec{v}} - \frac{e}{c}\vec{A}_{xc}(\vec{r})\frac{\vec{j}(\vec{r})}{n(\vec{r})} - \frac{e^2}{2mc^2}A_{xc}^2(\vec{r}) \tag{4.7}$$

$$\vec{A}_{eff}(\vec{r}) = \vec{A}(\vec{r}) + \vec{A}_{xc}(\vec{r}) = \vec{A}(\vec{r}) + \frac{1}{n(\vec{r})}\vec{\nabla} \times \frac{\delta \overline{E}_{xc}[n, \vec{\nu}]}{\delta \vec{\nu}(\vec{r})}\bigg|_n \tag{4.8}$$

From these equations, it is easily seen that both $\vec{A}_{xc}$ and V_{eff} are gauge invariant, and therefore Eq. (3.3) (or, equivalently, Eq. (3.9)) has the correct gauge-covariance of the usual Schroedinger equation. To show this, suppose we have found the self-consistent solution corresponding to a vector potential $\vec{A}(r)$. Let us multiply each of the self-consistent $\psi_i's$ by a phase factor $exp[\frac{ie}{\hbar c}\Lambda(\vec{r}_i)]$, so that the new paramagnetic current is given by Eq. (4.2). Let us insert the new $\vec{j}_p$ in $\vec{A}_{eff}$ and V_{eff} and let us also replace $\vec{A}$ according to Eq. (4.1). Then, due to the invariance of V_{eff} and $\vec{A}_{xc}$, one finds that the new $\psi_i's$ are indeed the self-consistent solution of the transformed equation. Thus, the proof of gauge-covariance is complete.

In addition, from the form of $\vec{A}_{xc}$, one derives the remarkable identity

$$\vec{\nabla}[n(\vec{r})\vec{A}_{xc}(\vec{r})] = 0$$

This, in combination with the usual property of the Schroedinger equation mentioned at the beginning of this section, guarantees that the continuity equation is automatically satisfied by any self-consistent solution of Eq. (3.3).

As a final point related to gauge-invariance, we briefly discuss the translational symmetry of the effective potentials in the practically important case of a uniform magnetic field and periodic crystal potential. It is well known that a translation in space by a lattice basis vector $\vec{a}$ is equivalent in this case to a gauge transformation[16], whose generating function is

$$\Lambda(\vec{r}) \sim (\vec{B} \times \vec{a}) \cdot \vec{r}$$

Therefore, the gauge-covariance of the VR formalism automatically ensures the periodicity of the density, current-density, effective scalar potential and exchange-correlation part of the effective vector potential. The only non-periodic quantity in the self-consistent equations is the external vector potential $\vec{A}$. This implies that all the usual theorems about the classification of the solutions of the one electron Schroedinger equation according to the representations of the magnetic translation group, are also good for the VR self-consistent equation. For instance, a Bloch-wave representation of the solutions to the VR equation is allowed whenever $\vec{B}$ is along one of the lattice basis vectors, and the number of magnetic flux quanta per unit cell is rational[16].

5. LOCAL DENSITY APPROXIMATION

Density functional theory owes most of its popularity to the existence of a simple but remarkably successful local density approximation[3] (LDA), which expresses the E_{xc} functional in terms of the exchange-correlation energy (per electron) $\epsilon_{xc}(n)$ of a uniform electron gas:

$$E_{xc}^{LDA}[n] = \int n(\vec{r})\epsilon_{xc}(n(\vec{r}))d\vec{r} \quad . \tag{5.1}$$

In this section we develop a similar local density approximation for current-density functional theory. From the results of the previous section it is evident that in this case the LDA must be in terms of $n(\vec{r})$ and $\vec{\nu}(\vec{r})$, and thus will have the form

$$\overline{E}_{xc}^{LDA}[n,\vec{\nu}] = \int n(\vec{r})\epsilon_{xc}(n(\vec{r}),\mid \vec{\nu}(\vec{r}) \mid)d\vec{r} \tag{5.2}$$

where $\epsilon_{xc}(n,\vec{\nu})$ is now the exchange-correlation energy of an electron gas with uniform density and unform $\vec{\nu}(\vec{r})$. The question arises: what is the physical meaning of a uniform $\vec{\nu}(\vec{r})$ distribution? The answer to this question is easily obtained by considering the effect of a uniform magnetic field on a uniform electron gas. If the electron gas is at rest in the laboratory frame then the current-density $\vec{j}(\vec{r})$ must vanish everywhere (a way to see this is to remember that a translation in real space is equivalent to a gauge transformation and hence leaves all physical quantities unaffected. Thus, the current density can at most be constant, and, in the rest frame of the system the constant must be zero). The vanishing of the current density implies the uniformity of $\vec{\nu}(\vec{r})$, which is given by

$$\vec{\nu}(\vec{r}) = \vec{\nu} = -\frac{e}{mc}\vec{B} \tag{5.3}$$

Thus, we find that $\epsilon_{xc}(n,v)$ is the exchange-correlation energy (per electron) of a uniform electron gas in a uniform field B given by Eq. (5.3). This is quite satisfactory from a physical point of view. It also sheds some light on the physical meaning of $\vec{\nu}(r)$: it is essentially the value of the magnetic field which would locally produce the assigned distribution of paramagnetic current, if the density were sufficiently uniform.

With Eqs. (5.2) and (5.3) we are now in a position to give explicit forms for the LDA in various physical regimes. Of course, the LDA is expected to work well only when the scale of variation of n and $\vec{\nu}$ is much larger than any microscopic length scale (interelectronic distance, radius of the cyclotron orbits).

(i) Weak field regime ($\mu B << kT << E_F$)

The LDA in this regime was first derived by Vignale and Rasolt[12,13]. The magnetic field and the temperature are both small compared to the Fermi energy, but the temperature is sufficiently large to suppress de Haas-van Alphen oscillations in the energy. In this regime, the ground-state energy (per unit volume) of a uniform electron gas is given by second-order perturbation theory:

$$E(n,B) = E(n,0) - \frac{1}{2}\chi(n)B^2 \tag{5.4}$$

where χ is the orbital magnetic susceptibility. Inserting Eq. (5.3) for B and using Eq. (5.2), we obtain

$$\overline{E}_{xc}^{LDA}[n,\vec{v}] \cong E_{xc}^{LDA}[n] + \int \left[\frac{mk_F}{24\pi^2}\left(\frac{\chi}{\chi_0} - 1\right)\right]_{n=n(\vec{r})} \mid \vec{v}(\vec{r}) \mid^2 d\vec{r} \tag{5.5}$$

where $\chi_0 = -(e^2 k_F/12\pi^2 mc^2)$ is the non-interacting orbital susceptibility (k_F is the Fermi momentum). This result was obtained by VR[12,13], using a completely different method which did not explicitly rely on $\vec{j} = 0$. The present derivation is much simpler, and physically more transparent. It explains, in particular, the a priori surprising fact that the coefficient of $\mid \vec{v} \mid^2$ is proportional to the susceptibility itself, rather than its inverse, as in the weak field regimes of DFT and $SDFT$[3].

The remaining problem is, of course, the calculation of χ. An early calculation of this quantity was done by Rajagopal and Jain[17], using the static Thomas-Fermi approximation. A more complete treatment has been recently provided by the present authors[18]. Our calculation parallels the calculation of Ma and Brueckner[19] of the exchange-correlation energy of a weakly inhomogeneous electron gas, but is applied to the current-current correlation function. This calculation is exact in the high-density limit. (For technical details we refer the reader to our original paper.[18]) The results for χ at metallic densities are tabulated in Table I. Note that the many-body renormalization is quite small, in comparison to that of the spin-susceptibility. Correspondingly, the current-dependent corrections to E_{xc} in this regime are expected to be very small.

(ii) de Haas-van Alphen regime ($kT << \mu B << E_F$)

The exchange-correlation energy in this regime is dominant by quantum oscillation of the form[20]

$$E_{xc}^{osc}(B) = (eB/c)^{5/2}[4\pi^4\hbar^{1/2}]^{-1}\left(\frac{1}{m^*} - \frac{1}{m}\right)$$
$$\sum_{\ell}(-1)^{\ell}\,\ell^{-5/2}cos\left[\frac{\ell\omega_0}{B} - \frac{\pi}{4}\right] \tag{5.6}$$

where m^* is the quasiparticle effective mass and $\omega_0 = \hbar c \pi k_F^2/e$. Then using Eqs. (5.2) and (5.3) we find

$$E_{xc}^{osc}[n,\vec{v}] \approx \int E_{xc}^{osc}\left(\mid \frac{mc}{e}\vec{v}(\vec{r}) \mid\right) d\vec{r} \tag{5.7}$$

Note that the *LDA* functional retains an oscillatory structure in $1/\mid \vec{v} \mid$, but unless the density is strictly uniform, the space integration over different periodicities leads to the appearance of an extra power of $\mid \vec{v} \mid$ which strongly suppresses the oscillations in the limit $\vec{v} \to 0$.

Table I.

Comparison of many-body effects in the orbital and spin susceptibilities of the uniform electron gas. The non-interacting susceptibilities are $\chi_0^{spin} = 2.59 \times 10^{-6}/r_s$ (in cgs units) and $\chi_0^{orb} = -\frac{1}{3}\chi_0^{spin}$. The value of the spin-susceptibility is taken from Table II of Ref. 22.

r_s	χ^{orb}/χ_0^{orb}	$\chi^{spin}/\chi_0^{spin}$	r_s	χ^{orb}/χ_0^{orb}	$\chi^{spin}/\chi_0^{spin}$
1	0.981	1.15	5	0.926	1.79
2	0.970	1.31	6	0.909	1.98
3	0.957	1.46	7	0.892	-
4	0.942	1.62	8	0.875	-

(iii) Strong field regime ($\mu B \sim E_F$)

Perhaps the most interesting applications of *CDFT* will come from the study of systems in extremely high magnetic fields. Eqs. (5.2) and (5.3) provide an attractively simple approximation for the strong field situation. Unfortunately, to date, no reliable expression for the exchange–correlation energy in a uniform magnetic field is available[23]. Many-body calculations of this quantity will be welcome.

6. CURRENT- AND SPIN - DENSITY FUNCTIONAL THEORY

The inclusion of spin in the *CDFT* formalism has recently been discussed by Vignale and Rasolt[13]. In the most general case, the basic variables for current and spin-density functional theory (*CSDFT*) are the spin-density matrix

$$n_{\alpha\beta}(\vec{r}) = < \psi_\beta^+(\vec{r})\psi_\alpha(\vec{r}) > = n_0(\vec{r})\delta_{\alpha\beta} + \sum_{\lambda=1}^{3} n_\lambda(\vec{r})\sigma_{\alpha\beta}^\lambda \tag{6.1}$$

(σ_λ are Pauli matrices), and the spin-current-density matrix

$$\vec{j}_{p\alpha\beta}(\vec{r}) = -\frac{i\hbar}{2m} < \psi^+_\beta(\vec{r})\vec{\nabla}\psi_\alpha(\vec{r}) - [\vec{\nabla}\psi^+_\beta(\vec{r})]\psi_\alpha(\vec{r}) >$$
$$= \vec{j}_{p0}(\vec{r})\delta_{\alpha\beta} + \sum_{\lambda=1}^{3} \vec{j}_{p\lambda}(\vec{r})\sigma^\lambda_{\alpha\beta} \tag{6.2}$$

There are basically two alternatives for the formulation of the theory:

(i) Total current formulation.

Since the physical Hamiltonian (with the additional $\vec{B}\cdot\vec{S}$ term describing the coupling of the magnetic field to the spin) depends only on the *total* current-density operator $\vec{j}_p$ (i.e., the trace of $\vec{j}_{p\alpha\beta}$) one can treat E, E_{xc}, etc., as functionals of $n_{\alpha\beta}(r)$ and $\vec{v}(\vec{r})$. The self-consistent equation takes the form

$$\sum_\beta \left\{ \frac{1}{2m}\left(-i\hbar\vec{\nabla} + \frac{e}{c}\vec{A}_{eff}(\vec{r})\right)^2 + V_{eff}(\vec{r}) \right\}_{\alpha\beta} \psi_\beta(\vec{r}) = \epsilon\psi_\alpha(\vec{r}) \tag{6.3}$$

where $\vec{A}_{eff}$ and V_{eff} are now matrices in spin space, given by Eqs. (4.8) where the only difference is that $\delta E_{xc}/\delta n$ is replaced by $\delta E_{xc}/\delta n_{\alpha\beta}$. Note that, in this formulation, the effective vector potential is still spin-independent (i.e., proportional to the unit matrix in spin space), like the physical one. It is evident, from our previous discussions, that Eq. (6.3) is gauge-covariant and satisfies current conservation. A local density approximation for E_{xc} is obtained—in principle—by considering the exchange-correlation energy of an electron gas with uniform spin density matrix $n_{\alpha\beta}$, in the presence of a uniform magnetic field B given by Eq. (5.2).

The total current formalism is promising from a practical point of view, because it provides the exact ground-state energy with minimum departure from the conventional spin-density functional formalism. On the other hand, the spin-currents generated by the self-consistent solution of the one-electron equations in this scheme are completely void of physical meaning. Only the total current has physical meaning.

(ii) Spin-current formulation

Alternatively, as shown by Vignale and Rasolt[13], one can construct a much more complete theory that yields both the spin-density matrix and the spin-current density matrix. To this end, one regards E_{xc} as a functional of $n_{\alpha\beta}$ and $\vec{j}_{p\alpha\beta}$. The self-consistent equations are still given by Eq.

(6.3) but $A_{xc\alpha\beta} = \delta E_{xc}/\delta \vec{j}_{p\beta\alpha}$ is now a nontrivial matrix in spin space. In order to implement this formulation it is necessary to extend the space of admissible Hamiltonians to include the possibility of a fictitious spin-dependent vector potential (of course, the physical Hamiltonian will still have a spin-independent vector potential). Perhaps, the most important consequence of this extension is that it exposes some additional internal symmetries of E_{xc}, which would not be accessible in the total current formulation. Specifically, it was shown in Ref. 13 that a combined rotation and gauge transformation of the basic variables

$$
\begin{aligned}
& n_0(\vec{r}) \to n_0(\vec{r}) \\
& n_\lambda(\vec{r}) \to \sum_{\mu=1}^{3} R_{\lambda\mu}(\vec{r}) n_\mu(\vec{r}) \\
& \vec{j}_{po}(\vec{r}) \to \vec{j}_{po}(\vec{r}) + \frac{e}{mc}\left[n(\vec{r})\vec{\nabla}\Lambda_0(\vec{r}) + \sum_{\lambda=1}^{3} n_\lambda(\vec{r})\vec{\nabla}\Lambda_\lambda(\vec{r})\right] \\
& \vec{j}_{p\lambda}(\vec{r}) \to \sum_{\mu=1}^{3} R_{\lambda\mu}(\vec{r})\left[\vec{j}_{p\mu}(\vec{r}) + \frac{e}{mc}(n_\mu(\vec{r})\vec{\nabla}\Lambda_0(\vec{r}) \right. \\
& \qquad \left. + n_0(\vec{r})\vec{\nabla}\Lambda_\mu(\vec{r}))\right]
\end{aligned}
\tag{6.4}
$$

where $\Lambda_{\alpha\beta}(\vec{r}) = \Lambda_0\delta_{\alpha\beta} + \sum_{\lambda=1}^{3} \vec{\Lambda}(\vec{r}) \cdot \vec{\sigma}_{\alpha\beta}$ is an $\vec{r}$-dependent spin-matrix and $R_{\lambda\mu}$ is the matrix corresponding to a rotation by $2e \mid \vec{\Lambda} \mid /\hbar c$ along the direction of $\vec{\Lambda}$, is a symmetry operation within the enlarged Hamiltonian space. The ground-state wavefunction transforms according to the rule

$$
\begin{aligned}
& \psi(r_1 \cdots r_N; \sigma_1 \cdots \sigma_N) \to \\
& exp\left[i\frac{e}{\hbar c}\sum_{i=1}^{N} \Lambda(\vec{r}_i, \sigma_i)\right]\psi(\vec{r}_1 \cdots \vec{r}_N; \sigma_1 \cdots \sigma_N)
\end{aligned}
\tag{6.5}
$$

Hence, the transformation laws of F and T_s are easily worked out, and by a straightforward generalization of the arguments leading to Eq. (4.5) one finds that E_{xc} is invariant under the transformation. As a consequence of this invariance, one can show that the generalized continuity equation for the spin-currents

$$
\vec{\nabla} \cdot \left[\vec{j}_{p\lambda}(\vec{r}) + \frac{e}{mc} S_\lambda(\vec{r})\vec{A}(\vec{r})\right] = \frac{e}{mc}\left(\vec{B}(\vec{r}) \times \vec{S}(\vec{r})\right)_\lambda \tag{6.6}
$$

is automatically satisfied. ($\vec{S}(\vec{r}) = (1/2)Tr\vec{\sigma}\hat{n}(\vec{r})$ is the spin density). This confirms the claim that the spin-currents obtained from the solution of the self-consistent equations are the correct ones.

Unfortunately, due to the complexity of the invariance group (6.4), it has not been possible so far to determine a simple functional form, comparable to Eq. (4.6) for E_{xc}. However, in the case of magnetic fields of constant direction, the structure of the group simplifies considerably, and one can easily prove that the functional has the form

$$E_{xc}[n_\uparrow, n_\downarrow, \vec{j}_{p\uparrow}, \vec{j}_{p\downarrow}] = \overline{E}_{xc}[n_\uparrow, n_\downarrow, \vec{\nu}_\uparrow, \vec{\nu}_\downarrow]$$
$$(\vec{\nu}_\sigma = \vec{\nabla} \times (\vec{j}_{p\sigma}/n_\sigma)) \quad . \tag{6.7}$$

A local density approximation can again be obtained in this case, by considering a fictitious uniform system with two different densities for spin up and spin down electrons, in the presence of two magnetic fields which couple only to the orbits of spin up and spin down electrons respectively. The magnitudes of these fictitious fields are determined along the lines of section 5, by requiring the current densities for each spin component to vanish everywhere. In the weak limit, and for not too large spin polarization, it is reasonable to assume that the two spin components contribute equally to the orbital susceptibility. A reasonable form of the E_{xc} functional is therefore

$$\overline{E}_{xc}^{LDA}[n_\uparrow, n_\downarrow, \vec{\nu}_\uparrow, \vec{\nu}_\downarrow] \simeq \overline{E}_{xc}^{LDA}[n_\uparrow, n_\downarrow]$$
$$+ \sum_\sigma \int \frac{m k_{F\sigma}}{48\pi^2} \left(\frac{\chi}{\chi_0} - 1 \right) \mid \nu_\sigma(\vec{r}) \mid^2 d\vec{r} \tag{6.8}$$

as proposed in Ref. 13. We refer to this paper for further technical details.

7. SUMMARY AND OPEN QUESTIONS

The *CSDFT* reviewed in this chapter is a rigorous formulation of a many-body problem of nonrelativistic interacting fermions in gauge fields. In addition to spin polarization, it includes for the first time the effect of orbital currents. A central result is the self-consistent Schroedinger-type equation, showing that the replacement of the external vector potential by an effective one (including exchange-correlation effects) must be done in the linear term, but not in the quadratic one. Although, at first sight, this appears to violate gauge invariance, and hence the continuity equation, a careful consideration of the transformation of the effective potentials, based on the symmetry of E_{xc}, reveals that there is no such violation.

We emphasize that the appearance of an exchange-correlation contribution to the vector potential has nothing to do with the fact circulating currents generate, according to Maxwell's equations, a "classical" contribution to the magnetic field. This effect is extremely small $[0(v^2/c^2)]$ and

could be accounted for by replacing the external $\vec{A}(\vec{r})$ by the self-consistent potential

$$\vec{A}_{sc}(\vec{r}) = \vec{A}(\vec{r}) + \frac{e}{c}\int \frac{\vec{j}(\vec{r}')}{\mid \vec{r} - \vec{r}' \mid} d\vec{r}'$$

The origin of $\vec{A}_{xc}(\vec{r})$, on the other hand, is purely quantum-mechanical, and does not depend on the electron charge. It would be present in a system of neutral particles (such as ^3He atoms) if the ground-state wavefunction carried a finite orbital current.

An interesting question that still needs clarification is the relation of the VR theory to the relativistic formulation of Rajagopal and Callaway[5]. In that formulation the Fermion current is given by

$$j^{\mu}(\vec{r}) = \overline{\psi}(\vec{r})\gamma^{\mu}\psi(\vec{r})$$

where $\psi(\vec{r})$ is a four-component Dirac spinor, and does not depend on the vector potential. Similarly, the current-field interaction does not contain the cumbersome A^2 term that was so problematic in the nonrelativistic formulation. However, the price paid for this simplification is that one must now work with four-component spinors, i.e., electrons and positrons on equal footing. When the nonrelativistic limit is taken and the Dirac spinors are folded back into two-component Pauli spinors (for example, by application of the Foldy-Wouthuysen transformation) the A^2 term will reappear, and the relativistic current will go over to the nonrelativistic one. It is then clear, by the arguments given in the introduction, that the minimization of the energy functional will have to be with respect to j_p. Thus, if the nonrelativistic limit is taken correctly the final Schroedinger type equation and the form of E_{xc} must reduce to the VR form. This limit has not been explicitly verified to date.

At the time of writing there is no application of the $CDFT$ to physical systems to be reported. However, a number of theoretical questions have already been addressed within the frame of $CDFT$. A particularly interesting question is the following. It is known that the quasiparticle Fermi surface in an inhomogeneous many-body systems, differs, in general, from the Fermi surface associated with the Kohn-Sham eigenvalues[24]. It is also known, however, that, in the presence of a small magnetic field the ground state energy of a crystal exhibit $dHvA$ oscillations whose periodicity (in 1/B) is determined by the cross sectional area of the *quasiparticle* Fermi surface in a plane normal to the field[21]. Since $CDFT$ is an exact theory of the ground state energy, it appears that, at least in principle, one should be able to extract the correct shape of the quasiparticle Fermi surface without having to resort to the many-body self-energy. This question has recently been posed by Serene[25]. He argues that the eigenvalue sum in

the VR expression for the ground-state energy oscillates with periodicity dictated by the Kohn-Sham Fermi surface, and therefore, the oscillatory part of E_{xc} must be equal to the difference between the quasiparticle and the Kohn-Sham oscillations. Thus, in principle, all the information about the true quasiparticle Fermi surface is contained in E_{xc}. Unfortunately, it appears that this information is well concealed, because a simple local density approximation is not sufficient to extract it. This is consistent with our observation in Section 6, that the LDA form of E_{xc} does have $dHvA$ oscillations, but they are expected to be of higher order in B. The conclusion is that the important information about the quasiparticle Fermi surface must be hidden in the non-local parts of the E_{xc} functional. Experience with ordinary DFT shows that such nonlocal contributions are usually extremely difficult to approximate.

We briefly mention two more recent developments of $CDFT$. Oliveira, Gross, and Kohn[26], have developed a density-functional theory for superconductors which is to the Bogoliubov-de Gennes theory[27] of inhomogeneous superconductivity what the Kohn-Sham theory is to the Hartree theory. Using the VR formalism, the authors have been able to incorporate a magnetic field in their formalism. This generalization should prove useful in the study of magnetic penetration phenomena in superconductors.

Very recently, T.K. Ng[28] has studied the time-dependent generalization of $CDFT$ for electronic systems in weak electromagnetic fields. He finds that in the dynamic situation and in the linear response regime, a gauge-invariant theory can be formulated in terms of $\vec{j}$. The structure of his LDA E_{xc} functional in the weak field regime is, however, different from the VR one, because he works in the limit $\omega > 0, q = 0$, whereas VR work in the limit $q > 0, \omega = 0$.

In conclusion, the self-consistent $CDFT$ for magnetic systems appears now to be on firm theoretical grounds. An open challenge is to develop useful approximate forms for E_{xc} based on many-body calculations of the correlation energy. Explicit calculations of the properties of physical systems are also needed to establish the power and limitations of this new method.

REFERENCES

1. P. Hohenberg and W. Kohn, Phys. Rev. **136** B846 (1964).
2. W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).
3. For a review, see W. Kohn and P. Vashishta, in *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N.H. March (Plenum, New York, 1983), Chap. 2.
4. A time-dependent formulation of DFT has been recently achieved by Erich Runge and E.K.U. Gross, Phys. Rev. Lett. **52**, 997 (1984); E.K.U. Gross and Walter Kohn, Phys. Rev. Lett. **55**, 2850 (1985).
5. A.K. Rajagopal and J. Callaway, Phys. Rev. B **7**, 1912 (1973).
6. O. Gunnarsson and B.I. Lundqvist, Phys. Rev. B. **13**, 4274 (1976).
7. U. von Barth and L. Hedin, J. Phys. C. **5**, 1629 (1972); D.M. Ceperley, Phys. Rev. B **18**, 3126 (1978); S.H. Vosko, L. Wilk, and M. Nusair, Can J. Phys. **58**, 1200 (1980); M. Rasolt, Phys. Rev. B. **16**, 3234 (1977); C.D. Hu and D.C. Langreth, Phys. Scr. **32**, 391-396 (1985).
8. *The quantum Hall Effect*, edited by Richard E. Prange and Steven M. Girvin (Springer-Verlag, New York, 1987).
9. B.I. Halperin, *Possible States of a Three-dimensional Electron Gas in a Strong Magnetic Field*, Proceedings of the 18th International Conference on Low Temperature Physics, Kyoto, 1987 (Japanese Journal of Applied Physics, **26** (1987), Suppl. 26-3, Part 3); Z. Tesanović and B.I. Halperin, Phys. Rev. B **36**, 6888 (1987).
10. D. Yoshioka and P.A. Lee. Phys. Rev. **B27** (1988) 4986.
11. C.S. Wang, D.R. Grempel, and R.E. Prange, Phys. Rev. B **28**, 4284 (1983).
12. G. Vignale and M. Rasolt, Phys. Rev. Lett. **59** 2360 (1987) .
13. G. Vignale and M. Rasolt, Phys. Rev. B **37**, 10685 (1988).
14. We assume, as customary, that our distributions belong to the class of distributions that can be physically realized in some external fields.
15. We are assuming here that the current and density distributions of interest are representable by some external potentials both in the interacting and in the noninteracting case.
16. L.D. Landau and E.M. Lifshitz, Course of Theoretical Physics, (Pergamon, New York, 1980), Vol. IX, Sec. 60.
17. A.K. Rajagopal and K.P. Jain, Phys. Rev. A. **5**, 1475 (1972).
18. G. Vignale, M. Rasolt, and D.J. W. Geldart, Phys. Rev. B. **37**, 2502 (1988).
19. Shang-keng Ma and Keith A. Brueckner, Phys. Rev. **165**, 18 (1968).
20. L.D. Landau and E.M. Lifschitz, Course of Theoretical Physics (Pergamon, New York, 1980), Vol. IX, Sec. 63.

21. J.M. Luttinger, Phys. Rev. **121** (1961) 1251; Yu. A. Bychkov and L.P. Gorkov, Zh. Eksp. Teor. Fiz. **41** (1961) 1592 [Soc. Phys. JETP **14** (1962) 1132].
22. N. Iwamoto and D. Pines, Phys. Rev. B **29**, 3924 (1984).
23. The exchange energy of the three-dimensional electron gas in a uniform magnetic field has been calculated by R.W. Danz and M.L. Glasser, Phys. Rev. **4**, 96 (1971). In two-dimensions, the calculation has been done by M.L. Glasser and N.J.M. Horing, Phys. Rev. **B31** 4603 (1985).
24. D. Mearns, Phys. Rev. B. **38**, 5906 (1988).
25. J.W. Serene, preprint.
26. L.N. Oliveira, E.K.U. Gross, and W. Kohn, Phys. Rev. Lett. **60**, 2430 (1988).
27. P.G. de Gennes, Superconductivity of Metals and Alloys (Benjamin, New York, 1966).
28. Tai Kai Ng, preprint.

TIME-DEPENDENT DENSITY-FUNCTIONAL THEORY

E.K.U. Gross
W. Kohn

Department of Physics
University of California
Santa Barbara, CA 93106

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

1. INTRODUCTION

Density functional theory for stationary states or ensembles is a formulation of many-body theory in terms of the particle density, $\rho(\mathbf{r})$. Based on the work of Hohenberg and Kohn[1] (HK) and Kohn and Sham[2] (KS), it is now a mature subject which has had many successful applications. In retrospect, the stationary Thomas-Fermi theory[3,4] and the Hartree[5] and Hartree-Fock-Slater[6] equations can be viewed as approximations of density functional theory.

Time-dependent density functional theory as a complete formalism[7] is of more recent origin, although a time-dependent version of Thomas-Fermi theory was proposed as early as 1933 by Bloch.[8] The time-dependent Thomas-Fermi model provides good first estimates for a variety of physical processes such as atomic[9–11] and nuclear[12,13] collisions, nuclear giant resonances,[14–16] atomic photoabsorption,[17–19] and plasmons in inhomogeneous media.[20–25] The time-dependent Hartree and Hartree-Fock equations[26] have been used extensively for more quantitative studies of nuclear dynamics[27,28] and atomic[29–32] as well as molecular[33] collision processes. However these methods neglect correlation effects. Inclusion of such effects would require the time-dependent analog of the stationary KS equations.

The first, and rather successful, steps towards a time-dependent KS theory were taken by Peuckert[34] and by Zangwill and Soven.[35] These authors treated the linear density response of rare-gas atoms to a time-dependent external potential as the response of non-interacting electrons to an effective time-dependent potential. In analogy to stationary KS theory, this effective potential was assumed to contain an exchange-correlation part, $v_{xc}(\mathbf{r}t)$, in addition to the time-dependent external and Hartree terms:

$$v_{eff}(\mathbf{r}t) = v(\mathbf{r}t) + \int \frac{\rho(\mathbf{r}'t)}{|\mathbf{r}-\mathbf{r}'|}\, d^3r' + v_{xc}(\mathbf{r}t).$$

Peuckert suggested an iterative scheme for the calculation of v_{xc}, while Zangwill and Soven adopted the functional form of the static exchange-correlation potential,

$$v_{xc}(\mathbf{r}t) = \frac{\delta E_{xc}[\rho(\mathbf{r}'t)]}{\delta \rho(\mathbf{r}t)},$$

where $E_{xc}[\rho]$ is the exchange-correlation energy functional of ordinary density functional theory. For this functional, Zangwill and Soven employed the local density approximation,

$$E_{xc}[\rho(\mathbf{r}'t)] = \int \epsilon_{xc}(\rho(\mathbf{r}'t))\rho(\mathbf{r}'t)d^3r',$$

where $\epsilon_{xc}(\rho)$ is the exchange-correlation energy per particle of a static uniform electron gas of density ρ. The static approximation is obviously valid only if the time-dependence of $\rho(\mathbf{r}t)$ is sufficiently slow. In practice, however, it gave quite good results even for the case of rather rapid time-dependence.

Significant steps towards a rigorous foundation of time-dependent density functional theory were taken by Deb and Ghosh[36–39] and by Bartolotti[40–43] who formulated and explored Hohenberg-Kohn and Kohn-Sham type theorems for the time-dependent density. Each of these derivations, however, was restricted to a rather narrow set of allowable time-dependent potentials (to potentials periodic in time in the theorems of Deb and Ghosh, and to adiabatic processes in the work of Bartolotti).

Finally, a general formulation covering essentially all time-dependent potentials of interest was given by Runge and Gross.[7] A novel feature of this formalism, not present in ground-state density functional theory, is the dependence of the respective density functionals on the initial (many-particle) state $\Psi(t_0)$. A detailed description of the time-dependent density-functional formalism will be presented in section

2. The central result is a set of time-dependent KS equations which are structurally similar to the time-dependent Hartree equations but include (in principle exactly) all many-body correlations through a local time-dependent exchange correlation potential.

To date, most applications of the formalism fall in the regime of linear response. In section 3, we shall describe the linear-response limit of time-dependent density functional theory along with applications to the photo-response of atoms, molecules and metallic surfaces.

Beyond the regime of linear response, the description of atomic and nuclear collision processes appears to be a promising field of application where the time-dependent KS scheme could serve as an economical alternative to time-dependent configuration-interaction calculations.[44–49] So far, only a simplified version of the time-dependent KS scheme has been implemented in this context.[50–53]

Another possible application beyond the regime of linear response is the calculation of atomic multiphoton ionization which, in the case of hydrogen, has recently been found[54,55] to exhibit chaotic behaviour. A full-scale numerical solution of the time-dependent Schrödinger equation for a hydrogen atom placed in strong time-dependent electric fields has recently been reported.[56] A time-dependent Hartree-Fock calculation has been achieved for the multiphoton ionization of helium.[57] For heavier atoms an analogous solution of the time-dependent Kohn-Sham equations offers itself as a promising application of time-dependent density functional theory.

2. FORMAL FRAMEWORK

2.1. One-to-one mapping between time-dependent potentials and time-dependent densities

Density functional theory is based on the existence of an exact mapping between densities and external potentials. In the ground-

state formalism,[1] the existence proof relies on the Rayleigh-Ritz minimum principle for the energy. Straightforward extension to the time-dependent domain is not possible since a minimum principle is not available in this case. The proof given by Runge and Gross[7] for time-dependent systems is based directly on the Schrödinger equation*

$$i\,\frac{\partial}{\partial t}\Psi(t) = \hat{H}(t)\Psi(t). \tag{1}$$

We shall investigate the densities $\rho(\mathbf{r}t)$ of electronic systems evolving from a fixed initial (many-particle) state

$$\Psi(t_0) = \Psi_0 \tag{2}$$

under the influence of different external potentials of the form

$$\hat{V}(t) = \sum_{\sigma=\uparrow\downarrow}\int d^3r\ \ \hat{\psi}^{+}_{\sigma}(\mathbf{r})v(\mathbf{r}t)\hat{\psi}_{\sigma}(\mathbf{r}). \tag{3}$$

In the following discussion, the initial time t_0 is assumed to be finite and the potentials are required to be expandable in a Taylor series about t_0:

$$v(\mathbf{r}t) = \sum_{j=0}^{\infty}\frac{1}{j!}\,v_j(\mathbf{r})(t-t_0)^j. \tag{4}$$

No further assumptions concerning the size of the radius of convergence are made. It is sufficient that the radius of convergence is greater than zero. The initial state Ψ_0 is not required to be the ground state or some other eigenstate of the initial potential $v(\mathbf{r}t_0) = v_0(\mathbf{r})$. This means that the case of sudden switching is included in the formalism. On the other hand, potentials that are switched-on adiabatically from $t_0 = -\infty$ are excluded by the condition (4).

* Atomic units are used throughout.

Besides an external potential of the form (3), the Hamiltonian in Eq. (1) contains the kinetic energy of the electrons and their mutual Coulomb repulsion:

$$\hat{H}(t) = \hat{T} + \hat{U} + \hat{V}(t) \tag{5}$$

with

$$\hat{T} = \sum_{\sigma=\uparrow\downarrow} \int d^3r \;\; \hat{\psi}_\sigma^+(\mathbf{r})\left(-\frac{\nabla^2}{2}\right) \hat{\psi}_\sigma(\mathbf{r}) \tag{6}$$

and

$$\hat{U} = \tfrac{1}{2}\sum_{\sigma,\tau} \int d^3r \int d^3r' \;\; \hat{\psi}_\sigma^+(\mathbf{r})\hat{\psi}_\tau^+(\mathbf{r}') \frac{1}{|\mathbf{r}-\mathbf{r}'|} \hat{\psi}_\tau(\mathbf{r}')\hat{\psi}_\sigma(\mathbf{r}). \tag{7}$$

The number of electrons, N, is fixed.

We shall demonstrate in the following that the densities $\rho(\mathbf{r}t)$ and $\rho'(\mathbf{r}t)$ evolving from a common initial state Ψ_0 under the influence of two potentials $v(\mathbf{r}t)$ and $v'(\mathbf{r}t)$ are always different provided that the potentials differ by more than a purely time-dependent (r-independent) function:†

$$v(\mathbf{r}t) \;\neq\; v'(\mathbf{r}t) + c(t). \tag{8}$$

Since both potentials can be expanded in a Taylor series and since they differ by more than a time-dependent function, some of the expansion coefficients $v_j(\mathbf{r})$ and $v_j'(\mathbf{r})$ (cf. Eq. (4)) must differ by more than

† If v and v' differ by a purely time-dependent function, the resulting wave functions $\Psi(t)$ and $\Psi'(t)$ differ by a purely time-dependent phase factor and, consequently, the resulting densities ρ and ρ' are identical. This trivial case is excluded by the condition (8), in analogy to the ground-state formalism where the potentials are required to differ by more than a constant.

a constant. Hence there exists some smallest integer $k \geq 0$ such that

$$v_k(\mathbf{r}) - v'_k(\mathbf{r}) = \frac{\partial^k}{\partial t^k} [v(\mathbf{r}t) - v'(\mathbf{r}t)]_{t=t_0} \neq \text{const.} \tag{9}$$

From this inequality we shall deduce in a first step that the current densities $\mathbf{j}(\mathbf{r}t)$ and $\mathbf{j}'(\mathbf{r}t)$ corresponding to $v(\mathbf{r}t)$ and $v'(\mathbf{r}t)$ are different. In a second step it will be demonstrated that the densities $\rho(\mathbf{r}t)$ and $\rho'(\mathbf{r}t)$ are different.

Since we consider, by construction, only states $\Psi(t)$ and $\Psi'(t)$ that evolve from the same initial state Ψ_0, the current densities $\mathbf{j}$ and $\mathbf{j}'$ as well as the densities ρ and ρ' are identical at the initial time t_0:

$$\mathbf{j}(\mathbf{r}t_0) = \mathbf{j}'(\mathbf{r}t_0) = < \Psi_0 | \hat{\mathbf{j}}(\mathbf{r}) | \Psi_0 > \equiv \mathbf{j}_0(\mathbf{r}), \tag{10}$$

$$\rho(\mathbf{r}t_0) = \rho'(\mathbf{r}t_0) = < \Psi_0 | \hat{\rho}(\mathbf{r}) | \Psi_0 > \equiv \rho_0(\mathbf{r}). \tag{11}$$

$\hat{\mathbf{j}}(\mathbf{r})$ denotes the paramagnetic current-density operator

$$\hat{\mathbf{j}}(\mathbf{r}) = -\frac{i}{2} \sum_\sigma \{\hat{\psi}^+_\sigma(\mathbf{r})[\nabla \hat{\psi}_\sigma(\mathbf{r})] - [\nabla \hat{\psi}^+_\sigma(\mathbf{r})]\hat{\psi}_\sigma(\mathbf{r})\}, \tag{12}$$

and $\hat{\rho}(\mathbf{r})$ is the usual density operator

$$\hat{\rho}(\mathbf{r}) = \sum_\sigma \hat{\psi}^+_\sigma(\mathbf{r})\hat{\psi}_\sigma(\mathbf{r}). \tag{13}$$

The time evolution of the current densities $\mathbf{j}$ and $\mathbf{j}'$ is most easily compared by means of the equations of motion

$$\frac{\partial}{\partial t}\mathbf{j}(\mathbf{r}t) = -i < \Psi(t) | [\hat{\mathbf{j}}(\mathbf{r}), \hat{H}(t)] | \Psi(t) > \tag{14}$$

$$\frac{\partial}{\partial t}\mathbf{j}'(\mathbf{r}t) = -i < \Psi'(t) | [\hat{\mathbf{j}}(\mathbf{r}), \hat{H}'(t)] | \Psi'(t) > . \tag{15}$$

Taking the difference of these two equations at the initial time t_0 one finds

$$\frac{\partial}{\partial t}[\mathbf{j}(\mathbf{r}t)-\mathbf{j}'(\mathbf{r}t)]_{t=t_0} = -i<\Psi_0|[\hat{\mathbf{j}}(\mathbf{r}),(\hat{H}(t_0)-\hat{H}'(t_0))]|\Psi_0> = -\rho_0(\mathbf{r})\nabla(v(\mathbf{r}t_0)-v'(\mathbf{r}t_0)). \tag{16}$$

If the potentials $v(\mathbf{r}t)$ and $v'(\mathbf{r}t)$ differ at $t = t_0$ (i.e., if Eq. (9) is satisfied for $k = 0$) then the right-hand side of Eq. (16) cannot vanish identically. Consequently $\mathbf{j}(\mathbf{r}t)$ and $\mathbf{j}'(\mathbf{r}t)$ will become different infinitesimally later than t_0. If the minimum integer k for which Eq. (9) holds is greater than zero, one has to evaluate the $(k+1)$th time-derivative of $\mathbf{j}$ and $\mathbf{j}'$ at t_0. To this end, the quantum mechanical equation of motion

$$\frac{d}{dt} < \Psi(t)|\hat{O}(t)|\Psi(t) > \; = \; < \Psi(t)|\frac{\partial}{\partial t}\hat{O}(t)-i[\hat{O}(t),\hat{H}(t)]|\Psi(t) >$$

is applied $(k+1)$ times; the first time with $\hat{O}_1(t) = \hat{\mathbf{j}}(\mathbf{r})$, the second time with $\hat{O}_2(t) = -i[\hat{\mathbf{j}}(\mathbf{r}),\hat{H}(t)]$, etc. After some straightforward algebra one obtains

$$\frac{\partial^{k+1}}{\partial t^{k+1}}[\mathbf{j}(\mathbf{r}t)-\mathbf{j}'(\mathbf{r}t)]_{t=t_0} = -\rho_0(\mathbf{r})\nabla\{\frac{\partial^k}{\partial t^k}[v(\mathbf{r}t)-v'(\mathbf{r}t)]_{t=t_0}\} \neq 0, \tag{17}$$

where the last inequality follows from (9). Again the inequality (17) implies that $\mathbf{j}(\mathbf{r}t)$ and $\mathbf{j}'(\mathbf{r}t)$ will become different infinitesimally later than t_0. This completes the proof for the current densities.

In order to prove that the densities $\rho(\mathbf{r}t)$ and $\rho'(\mathbf{r}t)$ are different, one employs the continuity equation:

$$\frac{\partial}{\partial t}[\rho(\mathbf{r}t)-\rho'(\mathbf{r}t)] = -\nabla\cdot[\mathbf{j}(\mathbf{r}t)-\mathbf{j}'(\mathbf{r}t)]. \tag{18}$$

Evaluating the $(k+1)$ th time-derivative of Eq. (18) at the initial time t_0, one finds by insertion of Eq. (17)

$$\frac{\partial^{k+2}}{\partial t^{k+2}}\,[\rho(\mathbf{r}t)-\rho'(\mathbf{r}t)]_{t=t_0} = \nabla\cdot[\rho_0(\mathbf{r})\nabla w_k(\mathbf{r})] \tag{19}$$

where

$$w_k(\mathbf{r}) \;\equiv\; \frac{\partial^k}{\partial t^k}[v(\mathbf{r}t)-v'(\mathbf{r}t)]_{t=t_0} = v_k(\mathbf{r})-v_k'(\mathbf{r}). \tag{20}$$

It remains to be shown that the right-hand side of Eq. (19) cannot vanish identically if the condition (9), $w_k(\mathbf{r}) \neq$ const, is satisfied. The proof is by *reductio ad absurdum*: Assume that the right-hand side of Eq. (19) vanishes and consider the integral

$$\int d^3r\;\rho_0(\mathbf{r})[\nabla w_k(\mathbf{r})]^2 \tag{21}$$

which, by means of Green's theorem, can be written as

$$= -\int d^3r\;w_k(\mathbf{r})\nabla\cdot[\rho_0(\mathbf{r})\nabla w_k(\mathbf{r})] + \oint \mathbf{dS}\cdot(\rho_0(\mathbf{r})w_k(\mathbf{r})\nabla w_k(\mathbf{r})). \tag{22}$$

As long as one deals with realistic, i.e. experimentally realizable potentials $v(\mathbf{r}t)$, the surface integral in (22) vanishes: Experimentally realizable potentials are always produced by <u>real</u> (this means, in particular, normalizable) external charge densities $\rho_{ext}(\mathbf{r}t)$ so that, in the Coulomb gauge,

$$v(\mathbf{r}t) \;=\; \int d^3r'\frac{\rho_{ext}(\mathbf{r}'t)}{|\mathbf{r}-\mathbf{r}'|}. \tag{23}$$

A Taylor expansion of this equation about $t = t_0$ demonstrates that the expansion coefficients $v_k(\mathbf{r})$, and hence the functions $w_k(\mathbf{r})$, fall

off at least as $1/r$ asymptotically so that the surface integral vanishes. (For more general potentials see remark (iii) below.) The first term of (22), on the other hand, is zero by assumption. Thus the complete integral (21) vanishes and, since the integrand is nonnegative, one concludes

$$\rho_0(\mathbf{r})[\nabla w_k(\mathbf{r})]^2 = 0$$

in contradiction to $w_k(\mathbf{r}) \neq$ const.* This completes the proof of the theorem. A few remarks are in order at this point:

i) By virtue of the 1-1 correspondence established above (for a given Ψ_0), the time-dependent density determines the external potential uniquely up to within an additive purely time-dependent function. The potential, on the other hand, uniquely determines the time-dependent wave function, which can therefore be considered as a functional of the time-dependent density,

$$\Psi(t) = \Psi[\rho](t), \tag{24}$$

where $\Psi[\rho]$ is unique up to within a purely time-dependent phase factor. As a consequence, the expectation value of any quantum mechanical operator $\hat{Q}(t)$ is a unique functional of the density,

$$Q(t) = < \Psi[\rho](t)|\hat{Q}(t)|\Psi[\rho](t) > \equiv Q[\rho](t); \tag{25}$$

the ambiguity in the phase of $\Psi[\rho]$ cancels out. As a particular example, the right-hand side of Eq. (14) can be considered as a density

* It is assumed here that $\rho_0(\mathbf{r})$ does not vanish on a set of positive measure. (Otherwise no contradiction is reached since $\rho_0(\mathbf{r})$ could be zero in exactly those regions of space where $w_k(\mathbf{r}) \neq$ const.) Densities that vanish on a set of positive measure correspond to potentials having infinite barriers. Such potentials have to be excluded from the ordinary ground-state theory as well (see, e.g., Ref. 58).

functional which depends parametrically on $\mathbf{r}$ and t:

$$\mathbf{P}[\rho](\mathbf{r}t) \equiv -i < \Psi[\rho](t)|[\hat{\mathbf{j}}(\mathbf{r}), \hat{H}(t)]|\Psi[\rho](t) > . \tag{26}$$

This implies that the time-dependent particle and current densities can always be calculated (in principle exactly) from a set of "hydrodynamical" equations:

$$\begin{aligned} \frac{\partial}{\partial t}\rho(\mathbf{r}t) &= -\nabla\cdot\mathbf{j}(\mathbf{r}t) \\ \frac{\partial}{\partial t}\mathbf{j}(\mathbf{r}t) &= \mathbf{P}[\rho](\mathbf{r}t). \end{aligned} \tag{27}$$

In practice, the functional $\mathbf{P}[\rho]$ is of course only approximately known.

ii) In addition to their dependence on the density, the functionals $Q[\rho]$, defined by Eq. (25), implicitly depend on the initial state Ψ_0. No such state dependence exists in the ground state formalism. Of course one would prefer to have functionals of the density alone rather than functionals of $\rho(\mathbf{r}t)$ and Ψ_0. It should be noted, however, that for a large class of systems, namely those where Ψ_0 is a non-degenerate gound state, $Q[\rho]$ is indeed a functional of the density alone. This is because any non-degenerate ground state Ψ_0 is a unique functional of its density $\rho_0(\mathbf{r})$ by virtue of the traditional HK theorem.[1] One has to emphasize that this class of systems also contains all cases of sudden switching where Ψ_0 is a ground state, but not the ground state of the initial potential $v(\mathbf{r}t_0)$.†

iii) It is essential for the proof of the 1-1 correspondence that the surface integral in (22) vanishes. As long as one considers only realistic physical potentials of the form (23), a vanishing surface integral is

† Whether an exclusive dependence on the density alone can be demonstrated for an even larger class of systems is an open question at present.

guaranteed. However, if one allows more general potentials, the surface integral does not necessarily vanish. Consider a given initial state Ψ_0 that leads to a certain asymptotic form of $\rho_0(\mathbf{r})$. Then one can always find potentials which increase sufficiently steeply in the asymptotic region such that the surface integral does not vanish. For such cases, the right-hand side of Eq. (19) can be zero, as was demonstrated by Xu and Rajagopal[59] (see also Dhara and Gosh[60]) with the explicit examples

$$\rho_0(\mathbf{r}) = A\, e^{-\lambda r}/(\lambda r)^2, \qquad w_k(\mathbf{r}) = B\, e^{\lambda r}, \qquad A, \lambda > 0 \tag{28}$$

and

$$\rho_0(\mathbf{r}) = \text{const}, \qquad w_k(\mathbf{r}) \quad \text{satisfying} \quad \nabla^2 w_k(\mathbf{r}) = 0. \tag{29}$$

If one intends to include potentials other than those given by Eq. (23), an additional condition must be imposed to ensure that the surface integral in (22) vanishes. For a given initial state Ψ_0, the allowable potentials must satisfy the condition*

$$\lim_{r\to\infty} \{r^2 \rho_0(\mathbf{r}) v_j(\mathbf{r}) \nabla v_j(\mathbf{r})\} = 0 \tag{30}$$

for all coefficients $v_j(\mathbf{r})$ of the Taylor expansion (4). Thus, the density functionals $Q[\rho]$ given by Eq. (25) not only depend explicitly on the initial state Ψ_0; the latter also determines the set of allowable potentials and hence the domain of densities for which the functionals $Q[\rho]$ are defined. One has to emphasize, however, that the cases excluded by the additional condition (30) are largely unphysical. The examples

* The condition $\rho_0 w_k \nabla w_k \to 0$, given by Dhara and Ghosh[60] for $r\to\infty$, is not sufficient to ensure a vanishing surface integral since it does not account for the fact that the spherical surface increment $\mathbf{dS}$ increases proportional to r^2.

(28) and (29) of Xu and Rajagopal, for instance, lead to an infinite† potential energy $\int \rho(\mathbf{r}t)v(\mathbf{r}t)d^3r$ per particle in the vicinity of $t = t_0$. It would be desirable to prove the 1-1 correspondence under a physical condition, such as the requirement of a finite potential energy per particle, rather than under the more restrictive mathematical condition (30). Whether such a proof can be constructed is currently an open question.

2.2. Variational principle

The solution of the time-dependent Schrödinger equation (1) with initial condition (2) corresponds to a stationary point of the quantum mechanical action integral

$$A = \int_{t_0}^{t_1} dt \; < \Psi(t)|i\frac{\partial}{\partial t} - \hat{H}(t)|\Psi(t) > \; . \tag{31}$$

Since there is a 1-1 mapping between time-dependent wave functions, $\Psi(t)$, and time-dependent densities, $\rho(\mathbf{r}t)$, the corresponding density functional

$$A[\rho] = \int_{t_0}^{t_1} dt \; < \Psi[\rho](t)|i\frac{\partial}{\partial t} - \hat{H}(t)|\Psi[\rho](t) > \tag{32}$$

must have a stationary point at the correct time-dependent density (corresponding to the Hamiltonian $\hat{H}(t)$ and the initial state Ψ_0). Thus the correct density can be obtained by solving the Euler equation

$$\frac{\delta A[\rho]}{\delta \rho(\mathbf{r}t)} = 0 \tag{33}$$

with appropriate boundary conditions. The functional $A[\rho]$ can be

† Note that a function satisfying $\nabla^2 w_k(\mathbf{r}) = 0$ for all $\mathbf{r}$ cannot vanish asymptotically (except in the trivial case $w_k(\mathbf{r}) \equiv 0$), nor can it approach a constant asymptotic value (except in the case $w_k(\mathbf{r}) \equiv$ const).

written as

$$A[\rho] = B[\rho] - \int_{t_0}^{t_1} dt \int d^3r \; \rho(\mathbf{r}t) v(\mathbf{r}t) \tag{34}$$

with a universal (Ψ_0-dependent) functional $B[\rho]$, formally defined as

$$B[\rho] = \int_{t_0}^{t_1} dt \; < \Psi[\rho](t) | i\frac{\partial}{\partial t} - \hat{T} - \hat{U} | \Psi[\rho](t) > . \tag{35}$$

2.3. Time-dependent Kohn-Sham scheme

A particularly important application of the stationary action principle described in the last section is the derivation of a time-dependent KS scheme. To this end we first consider a system of non-interacting fermions subject to an external potential $v(\mathbf{r}t)$. The exact many-particle wave function is then a time-dependent Slater determinant $\Phi(t)$ which, by virtue of the Runge-Gross theorem, is a functional, $\Phi[\rho](t)$, of the time-dependent density. (Note that the argument of section 2.1 holds true for any particle-particle interaction U, in particular also for $U \equiv 0$. The functionals $\mathbf{P}[\rho]$, $B[\rho]$, etc., however depend on the particular U considered.) The action functional for non-interacting particles then has the form

$$A_s[\rho] = B_s[\rho] - \int_{t_0}^{t_1} dt \int d^3r \; \rho(\mathbf{r}t) v(\mathbf{r}t) \tag{36}$$

where

$$B_s[\rho] = \int_{t_0}^{t_1} dt \; < \Phi[\rho](t) | i\frac{\partial}{\partial t} - \hat{T} | \Phi[\rho](t) > . \tag{37}$$

According to the variational principle of section 2.2, the exact density can be obtained by solving the Euler equation

$$\frac{\delta A_s[\rho]}{\delta \rho(\mathbf{r}t)} = 0 = \frac{\delta B_s[\rho]}{\delta \rho(\mathbf{r}t)} - v(\mathbf{r}t). \tag{38}$$

Next we assume† that a single-particle potential $v_s(\mathbf{r}t)$ exists such that the interacting density is identical with the density

$$\rho(\mathbf{r}t) = \sum_{n=1}^{N} |\varphi_n(\mathbf{r}t)|^2 \tag{39}$$

of the non-interacting system described by the Schrödinger equation

$$i\frac{\partial}{\partial t}\,\varphi_n(\mathbf{r}t) = \left(-\frac{\nabla^2}{2} + v_s(\mathbf{r}t)\right)\varphi_n(\mathbf{r}t). \tag{40}$$

If such a potential exists it must be unique (by virtue of the Runge-Gross theorem) and, in view of Eq. (38), it can be represented as

$$v_s(\mathbf{r}t) = \frac{\delta B_s[\xi]}{\delta \xi(\mathbf{r}t)} \Big|_{\xi(\mathbf{r}t)=\rho(\mathbf{r}t)}, \tag{41}$$

where the functional derivative of B_s has to be evaluated at the exact interacting (= non-interacting) density $\rho(\mathbf{r}t)$.

To determine the potential $v_s(\mathbf{r}t)$ more explicitly one employs the variational principle for the interacting system. First one rewrites the total action functional (34) as

$$A[\rho] = B_s[\rho] - \int_{t_0}^{t_1} dt \int d^3r\, \rho(\mathbf{r}t) v(\mathbf{r}t) - \tfrac{1}{2}\int_{t_0}^{t_1} dt \int d^3r \int d^3r'\, \frac{\rho(\mathbf{r}t)\rho(\mathbf{r}'t)}{|\mathbf{r}-\mathbf{r}'|} - A_{xc}[\rho], \tag{42}$$

where the "exchange-correlation" part of the action functional is formally defined as

† This assumption is analogous to the familiar assumption of simultaneous interacting and non-interacting v-representability made in the ground-state formalism.

$$A_{xc}[\rho] = B_s[\rho] - \tfrac{1}{2}\int_{t_0}^{t_1} dt \int d^3r \int d^3r' \, \frac{\rho(\mathbf{r}t)\rho(\mathbf{r}'t)}{|\mathbf{r}-\mathbf{r}'|} - B[\rho]. \tag{43}$$

Inserting (42) in the variational equation (33) one finds

$$\frac{\delta B_s[\rho]}{\delta \rho(\mathbf{r}t)} - [v(\mathbf{r}t) + \int d^3r' \, \frac{\rho(\mathbf{r}'t)}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta A_{xc}[\rho]}{\delta \rho(\mathbf{r}t)}] = 0. \tag{44}$$

Since this equation is solved only by the exact interacting density one obtains, by comparison with Eq. (41), the explicit representation

$$v_s(\mathbf{r}t) = v(\mathbf{r}t) + \int d^3r' \, \frac{\rho(\mathbf{r}'t)}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta A_{xc}[\rho]}{\delta \rho(\mathbf{r}t)}. \tag{45}$$

Eqs. (39), (40) and (45) constitute the time-dependent KS scheme. Compared to the time-dependent Hartree-fock method, this scheme has two advantages:

a) The time-dependent effective potential $v_s(\mathbf{r}t)$ is local (i.e. a multiplication operator in real space).

b) Correlation effects are included.

Explicit approximations for the time-dependent exchange-correlation potential

$$v_{xc}(\mathbf{r}t) = \frac{\delta A_{xc}[\rho]}{\delta \rho(\mathbf{r}t)} \tag{46}$$

will be discussed in section 3.2 in the context of linear response theory.

The density functionals $P[\rho], B[\rho], A_{xc}[\rho]$, etc, are well-defined only for "v-representable" densities, i.e., for densities that come from some time-dependent potential satisfying Eqs. (4) and (30). Since the restriction to v-representable densities is difficult to implement in the basic variational principle of section 2.2, a Levy-Lieb-type[58,61,62] extension of the respective functionals to arbitrary (non-negative, normalizable) functions $\rho(\mathbf{r}t)$ appears desirable. Two different proposals

of this type have been put forward so far.[63,64]

Besides these mathematical generalizations, a number of extensions of the time-dependent density-functional formalism to physically different situations have been developed. Those include spin-polarized systems,[65] multicomponent systems,[66] time-dependent ensembles,[67,68] and external vector potentials.[64,69]

3. FREQUENCY-DEPENDENT LINEAR RESPONSE

3.1. Selfconsistent equations for the linear density response

In this section we shall consider the ground-state response of electronic systems. In that case, the initial density $\rho_0(\mathbf{r})$ resulting from the potential $v_0(\mathbf{r})$ can be calculated from the ordinary (ground-state) KS scheme:

$$(-\frac{\nabla^2}{2}+v_0(\mathbf{r})+\int d^3r' \frac{\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}+v_{xc}[\rho_0](\mathbf{r}))\varphi_i^{(0)}(\mathbf{r}) = \epsilon_i \varphi_i^{(0)}(\mathbf{r}) \quad (47)$$

$$\rho_0(\mathbf{r}) = \sum_{i(occ)} |\varphi_i^{(0)}(\mathbf{r})|^2. \quad (48)$$

At $t = t_0$ a time-dependent perturbation $v_1(\mathbf{r}t)$ is switched on, i.e. the total external potential is given by

$$v(\mathbf{r}t) = \begin{cases} v_0(\mathbf{r}) & \text{for } t < t_0 \\ v_0(\mathbf{r}) + v_1(\mathbf{r}t) & \text{for } t \geq t_0. \end{cases} \quad (49)$$

Our goal is to calculate the linear density response $\rho_1(\mathbf{r}t)$ which is conventionally expressed in terms of the full response function χ as

$$\rho_1(\mathbf{r}t) = \int d^3r' \int_{t_0}^{\infty} dt' \chi(\mathbf{r}t, \mathbf{r}'t') v_1(\mathbf{r}'t'). \quad (50)$$

Since the time-dependent KS equations (39), (40), (45) provide a formally exact way of calculating the time-dependent density, the exact

linear density response of the interacting system can alternatively be calculated as the density response of the non-interacting KS system:

$$\rho_1(\mathbf{r}t) = \int d^3r' \int_{t_0}^{\infty} dt' \chi_{KS}(\mathbf{r}t, \mathbf{r}'t') v_s^{(1)}(\mathbf{r}'t'). \tag{51}$$

$v_s^{(1)}(\mathbf{r}t)$ is the time-dependent KS potential (45), calculated to first order in the perturbing potential. It can be written in the form

$$v_s^{(1)}(\mathbf{r}t) = v_1(\mathbf{r}t) + \int d^3r' \frac{\rho_1(\mathbf{r}'t)}{|\mathbf{r} - \mathbf{r}'|} + \int d^3r' \int dt' f_{xc}(\mathbf{r}t, \mathbf{r}'t') \rho_1(\mathbf{r}'t'). \tag{52}$$

f_{xc} denotes the exchange-correlation response kernel which is formally defined as the functional derivative of the time-dependent exchange-correlation potential (46),

$$f_{xc}(\mathbf{r}t, \mathbf{r}'t') = \frac{\delta v_{xc}[\rho](\mathbf{r}t)}{\delta \rho(\mathbf{r}'t')} \Big| \rho = \rho_0, \tag{53}$$

evaluated at the initial ground-state density $\rho_0(\mathbf{r})$.

Eqs. (51) and (52) constitute the KS equations for the linear density response. Given some approximation for the exchange-correlation kernel f_{xc}, these equations provide a selfconsistent scheme to calculate the density response $\rho_1(\mathbf{r}t)$.

The response function χ_{KS} is relatively easy to calculate[35,70] (as opposed to the full response function χ). In terms of the stationary KS orbitals of Eq. (47), the Fourier transform of $\chi_{KS}(\mathbf{r}t, \mathbf{r}'t')$ with respect to $(t-t')$ is given by

$$\chi_{KS}(\mathbf{r}, \mathbf{r}'; \omega) = \sum_{k,j} (f_k - f_j) \frac{\varphi_k^{(0)}(\mathbf{r})^* \varphi_j^{(0)}(\mathbf{r}) \varphi_j^{(0)}(\mathbf{r}')^* \varphi_k^{(0)}(\mathbf{r}')}{\omega - (\epsilon_j - \epsilon_k) + i\delta} \tag{54}$$

where f_k and f_j are Fermi occupation factors. The summations in

(54) run over all KS orbitals (including the continuum states).

For the purpose of constructing approximations to the exchange-correlation kernel f_{xc}, it is useful to express this quantity in terms of the full response function χ. A formally exact representation of f_{xc} is readily obtained by solving Eq. (50) for v_1 and inserting the result in (52). Eq. (51) then yields

$$f_{xc}(\mathbf{r}t,\mathbf{r}'t') = \chi_{KS}^{-1}(\mathbf{r}t,\mathbf{r}'t') - \chi^{-1}(\mathbf{r}t,\mathbf{r}'t') - \frac{\delta(t-t')}{|r-r'|}. \tag{55}$$

By virtue of the Runge-Gross theorem, two potentials

$$v(\mathbf{r}t) = v_0(\mathbf{r}) + v_1(\mathbf{r}t) \quad \text{and} \quad v'(\mathbf{r}t) = v_0(\mathbf{r}) + v_1'(\mathbf{r}t)$$

always lead to different densities $\rho(\mathbf{r}t)$ and $\rho'(\mathbf{r}t)$ provided that $v_1(\mathbf{r}t) \neq v_1'(\mathbf{r}t) + c(t)$. This statement does not automatically imply that the linear density responses $\rho_1(\mathbf{r}t)$ and $\rho_1'(\mathbf{r}t)$ produced by $v_1(\mathbf{r}t)$ and $v_1'(\mathbf{r}t)$ are different; the difference between $\rho(\mathbf{r}t)$ and $\rho'(\mathbf{r}t)$ might show up only in higher perturbative orders. However, since the right-hand side of Eq. (19) is manifestly of first order in the perturbation, $\rho_1(\mathbf{r}t)$ and $\rho_1'(\mathbf{r}t)$ must in fact be different. Therefore the time-dependent KS scheme for the linear density response (Eqs. (51) and (52)) is well-defined and the inverse response operators χ^{-1} and χ_{KS}^{-1} in Eq. (55) must exist[71] on the set of v-representable densities that correspond to the set of external potentials specified in Sec. 2.1. On a larger set of potentials, the response functions χ and χ_{KS} are not necessarily invertible. In particular, the frequency-dependent response operators $\chi(\mathbf{r},\mathbf{r}';\omega)$ and $\chi_{KS}(\mathbf{r},\mathbf{r}';\omega)$ cannot always be inverted since they determine, by definition, the linear density response to adiabatically switched monochromatic perturbations,

$$v_1(\mathbf{r}t) = v_1(\mathbf{r}\omega)\, e^{-i\omega t}\, e^{-\eta|t|}, \quad -\infty \leq t \leq \infty,$$

which are not covered by the Runge-Gross theorem (due to the condition (4)). As a matter of fact, it has been demonstrated[72,73] that the response function $\chi(\mathbf{r},\mathbf{r}';\omega)$ of suitably constructed non-interacting systems can have eigenvectors with vanishing eigenvalues at isolated frequencies ω_0. Consequently the adiabatic density response vanishes at these frequencies and $\chi(\mathbf{r},\mathbf{r}';\omega_0)$ is not invertible.

We conclude this section with a discussion of the homogeneous electron gas, perturbed by an adiabatically switched potential consisting of a single Fourier component in $(q\omega)$-space:

$$v_1(\mathbf{r}t) = e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} v_1(\mathbf{q}\omega)\, e^{-\eta|t|}, \quad -\infty \leq t \leq \infty.$$

Again the Runge-Gross theorem is not applicable in this case. However, for this particular system the exchange-correlation kernel can be constructed in explicit terms: The linear density response $\rho_1(\mathbf{q}\omega)$ resulting from the perturbation $v_1(\mathbf{q}\omega)$ is determined by the full response function $\chi^{hom}(\mathbf{q}\omega)$ of the homogeneous electron gas,

$$\rho_1(\mathbf{q}\omega) = \chi^{hom}(q\omega) v_1(\mathbf{q}\omega). \tag{56}$$

Defining

$$f_{xc}^{hom}(q\omega) \equiv \frac{1}{\chi_0(q\omega)} - \frac{1}{\chi^{hom}(q\omega)} - \frac{4\pi}{q^2} \tag{57}$$

with χ_0 being the Lindhard function, one finds by insertion in (56):

$$\rho_1(\mathbf{q}\omega) = \chi_0(q\omega)\{v_1(\mathbf{q}\omega) + \frac{4\pi}{q^2}\rho_1(\mathbf{q}\omega) + f_{xc}^{hom}(q\omega)\rho_1(\mathbf{q}\omega)\}. \tag{58}$$

Comparison of Eq. (58) with Eqs. (51)–(52) shows that the Fourier transform $f_{xc}^{hom}(\mathbf{r}-\mathbf{r}', t-t')$ of $f_{xc}^{hom}(q\omega)$ is indeed the exchange-

correlation kernel of the homogeneous electron gas. $f_{xc}^{hom}(q\omega)$ is closely related to the so-called local field correction $G(q\omega)$ of the homogeneous electron gas:

$$f_{xc}^{hom}(q\omega) = -\frac{4\pi}{q^2} G(q\omega). \tag{59}$$

Using well-known features of the electron gas response functions, f_{xc}^{hom} can be shown to have the following exact properties:

1. As a consequence of the compressibility sum rule one finds[74]

$$\lim_{q\to 0} f_{xc}^{hom}(q, \omega = 0) = \frac{d^2}{d\rho^2}(\rho\epsilon_{xc}(\rho)) \equiv f_0(\rho), \tag{60}$$

where $\epsilon_{xc}(\rho)$ denotes the exchange-correlation energy per particle of the homogeneous electron gas.

2. The third-frequency-moment sum rule leads to[75]

$$\begin{aligned} &\lim_{q\to 0} f_{xc}^{hom}(q, \omega = \infty) \\ &= -\frac{4}{5}\rho^{2/3}\frac{d}{d\rho}\left(\frac{\epsilon_{xc}(\rho)}{\rho^{2/3}}\right) + 6\rho^{1/3}\frac{d}{d\rho}\left(\frac{\epsilon_{xc}(\rho)}{\rho^{1/3}}\right) \equiv f_\infty(\rho). \end{aligned} \tag{61}$$

3. According to the best estimates[76,77] of $\epsilon_{xc}(\rho)$, the following relation holds for all densities

$$f_0(\rho) < f_\infty(\rho) < 0. \tag{62}$$

4. In the static limit ($\omega = 0$), the short-wavelength behavior is given by[78]

$$\lim_{q\to\infty} f_{xc}^{hom}(q,\omega=0) \;=\; -\frac{4\pi}{q^2}(1-g(0)), \tag{63}$$

where $g(r)$ denotes the pair correlation function.

5. For frequencies $\omega\neq 0$, the short-wavelength behavior is[79]

$$\lim_{q\to\infty} f_{xc}^{hom}(q,\omega\neq 0) \;=\; -\frac{2}{3}\cdot\frac{4\pi}{q^2}(1-g(0)). \tag{64}$$

6. $f_{xc}^{hom}(q\omega)$ is a complex-valued function satisfying the symmetry relations

$$Re\, f_{xc}^{hom}(q,\omega) = Re\, f_{xc}^{hom}(q,-\omega) \tag{65}$$

$$Im\, f_{xc}^{hom}(q,\omega) = -Im\, f_{xc}^{hom}(q,-\omega). \tag{66}$$

7. $f_{xc}^{hom}(q\omega)$ is an analytic function of ω in the upper half of the complex ω-plane and approaches a real function $f_\infty(q)$ for $\omega\to\infty$.[80] Therefore, the function $f_{xc}^{hom}(q\omega)-f_\infty(q)$ satisfies standard Kramers-Kronig relations:

$$Re\, f_{xc}^{hom}(q\omega) - f_\infty(q) = \mathbf{P}\int_{-\infty}^{\infty}\frac{d\omega'}{\pi}\,\frac{Im\, f_{xc}^{hom}(q\omega')}{\omega'-\omega} \tag{67}$$

$$Im\, f_{xc}^{hom}(q\omega) = -\mathbf{P}\int_{-\infty}^{\infty}\frac{d\omega'}{\pi}\,\frac{Re\, f_{xc}^{hom}(q\omega') - f_\infty(q)}{\omega'-\omega}. \tag{68}$$

8. The imaginary part of f_{xc}^{hom} exhibits the high-frequency behavior

$$\lim_{\omega\to\infty} Im\, f_{xc}^{hom}(q,\omega) \;=\; -\frac{c}{\omega^{3/2}}. \tag{69}$$

for any $q < \infty$.[81] A second-order perturbation expansion[81,82] of the irreducible polarization propagator leads to the high-density limit

$$c = 23\pi/15. \tag{70}$$

9. In the same limit, the real part of f_{xc}^{hom} behaves like[83]

$$\lim_{\omega\to\infty} Re\, f_{xc}^{hom}(q,\omega) = f_\infty(q) + \frac{c}{\omega^{3/2}}. \tag{71}$$

Since $c > 0$, the infinite-frequency value f_∞ is approached from above. This implies, in view of the relation (62), that $Re\, f_{xc}^{hom}(q=0,\omega)$ cannot grow monotonically from f_0 to f_∞.

The above features of f_{xc}^{hom} are valid for a three-dimensional electron gas. Analogous results have been obtained for the two-dimensional case.[81,84,85]

3.2. Approximations for the exchange-correlation kernel

The time-dependent density-functional formalism described in section 2 can be shown[51] to reduce to the traditional HKS formalism for the class of static initial value problems in which the potential is time-independent, $v(\mathbf{r}t) = v(\mathbf{r})$, and where the initial state Ψ_0 is the ground state of $v(\mathbf{r})$. In particular, the functional $A_{xc}[\rho]$ reduces to the usual exchange-correlation energy if ground-state densities are inserted for $\rho(\mathbf{r}t)$:

$$A_{xc}[\bar\rho] = (t_1 - t_0)\cdot E_{xc}[\bar\rho] \quad \text{for all ground-state densities } \bar\rho(r). \tag{72}$$

This has an immediate consequence for possible approximations to the functional $A_{xc}[\rho]$: If an approximate expression for $A_{xc}[\rho]$ both satisfies (72) and is local in time, it must necessarily have the form

$$\int_{t_0}^{t_1} dt\; E_{xc}[\rho]\Big|_{\rho=\rho(\mathbf{r}t)}. \tag{73}$$

Of particular interest is the adiabatic local density approximation

$$A_{xc}^{ALDA}[\rho] \equiv \int_{t_0}^{t_1} dt \int d^3r \ [\rho \cdot \epsilon_{xc}(\rho)]\Big|_{\rho=\rho(\mathbf{r}t)} \tag{74}$$

which is local in time and space. By Eqs. (46) and (53), the exchange-correlation kernel corresponding to (74) takes the form

$$f_{xc}^{ALDA}(\mathbf{r}t, \mathbf{r}'t') = \delta(t-t')\delta(\mathbf{r}-\mathbf{r}')\frac{d^2}{d\rho^2}[\rho \cdot \epsilon_{xc}(\rho)]\Big|_{\rho=\rho_0(\mathbf{r})}. \tag{75}$$

Obviously, the static exchange-correlation energy can be expected to be a good approximation only for very slow time-dependent processes, i.e., one implicitly invokes an adiabatic approximation. This is also reflected in the fact that (75) is identical with the zero-frequency limit (60) of the homogeneous exchange-correlation kernel:

$$f_{xc}^{ALDA}(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r}-\mathbf{r}') f_{xc}^{hom}(q=0, \omega=0; \rho)\Big|_{\rho=\rho_0(\mathbf{r})}. \tag{76}$$

Going beyond the adiabatic local density approximation, Gross and Kohn[83] have proposed an approximation for the exchange-correlation kernel which explicitly incorporates the frequency dependence. With respect to spatial coordinates, a local approximation as in (76) is employed:

$$f_{xc}^{LDA}(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r}-\mathbf{r}') f_{xc}^{hom}(q=0, \omega; \rho)\Big|_{\rho=\rho_0(\mathbf{r})}. \tag{77}$$

The frequency-dependence of $Im\, f_{xc}^{hom}(q=0,\omega)$ is approximated by a Padé-like interpolation between the high- and the low-frequency limits:

$$Im\, f_{xc}^{hom}(q=0,\omega) = \frac{a(\rho)\cdot\omega}{(1+b(\rho)\cdot\omega^2)^{5/4}} \tag{78}$$

where

$$a(\rho) = -c(\gamma/c)^{5/3}(f_\infty(\rho)-f_0(\rho))^{5/3}$$

$$b(\rho) = (\gamma/c)^{4/3}(f_\infty(\rho)-f_0(\rho))^{4/3}$$

$$\gamma = \frac{[\Gamma(1/4)]^2}{4\sqrt{2\pi}}.$$

f_0, f_∞, and c are given by Eqs. (60), (61), and (70), respectively. For the exchange-correlation energy $\epsilon_{xc}(\rho)$ (which enters f_0 and f_∞), the parametrization of Vosko, Wilk and Nusair[77] is employed. Using the Kramers-Kronig relation (67), the real part can be expressed as

$$\begin{aligned} &Re\, f_{xc}^{hom}(q=0,\omega) \\ &= f_\infty + \frac{a}{\pi s^2}\cdot\sqrt{\frac{8}{b}}\cdot[2E(1/\sqrt{2}) - \frac{1+s}{2}\Pi(\frac{1-s}{2},\frac{1}{\sqrt{2}}) \\ &- \frac{1-s}{2}\Pi(\frac{1+s}{2},\frac{1}{\sqrt{2}})], \quad s^2 = 1+b\omega^2. \end{aligned} \tag{79}$$

E and Π are complete elliptic integrals of the second and third kind in the standard notation of Byrd and Friedman.[86]

Figs. 1 and 2 show the real and imaginary part of f_{xc}^{hom} as calculated from (78) and (79). The functions are plotted for the two density values corresponding to $r_s = 2$ and $r_s = 4$. For the lower density value ($r_s = 4$), a considerable frequency-dependence is found. The dependence on ω becomes less pronounced for higher densities. In the extreme high-density limit, the difference between f_0 and f_∞ tends to zero. One finds the exact result

$$f_\infty - f_0 \sim r_s^2 \qquad \text{for } r_s \to 0. \tag{80}$$

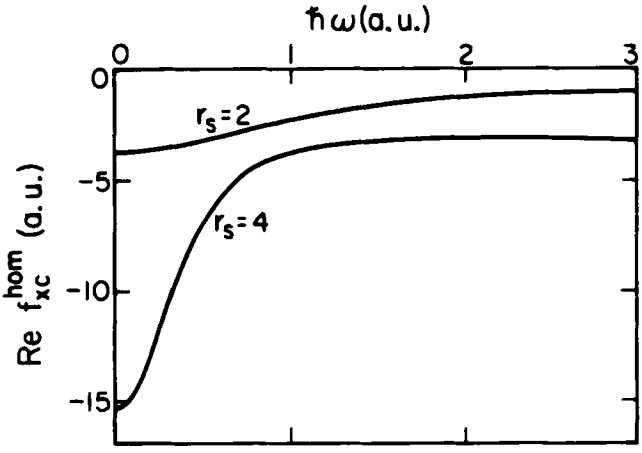


Fig. 1. Real part of the parametrization for $f_{xc}^{hom}(q=0,\omega)$ (from Ref. 75).

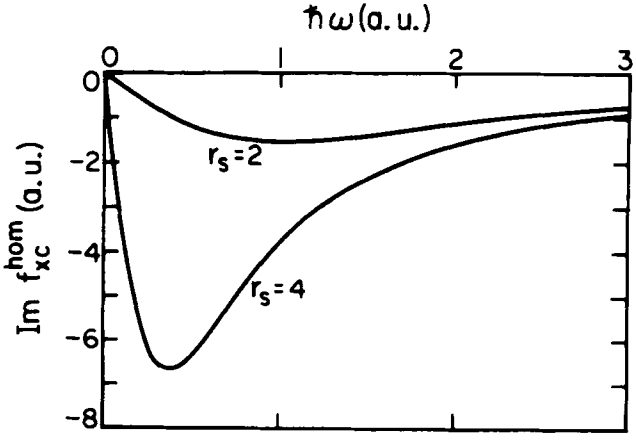


Fig. 2. Imaginary part of the parametrization for $f_{xc}^{hom}(q=0,\omega)$ (from Ref. 75).

At the same time, the depth of the minimum of Imf_{xc}^{hom} decreases, again proportional to r_s^2.

We finally mention that an extension of the parametrization (78) to nonvanishing q was given by Dabrowski.[87] A similar interpolation for the exchange-correlation kernel of the 2-dimensional electron gas has been derived by Holas and Singwi.[81]

3.3. Applications

The time-dependent KS scheme defined by Eqs. (51) and (52) has been successfully applied to the photo-response of atoms[35,88,89] and molecules,[90,91] metallic[92–98] and semiconductor[99] surfaces and, most recently, of bulk semiconductors.[100]

In the present context, we restrict ourselves to sufficiently low radiation frequencies such that the wavelength is much larger than the atomic radius and that Compton scattering is negligible. In the case of atoms, this condition is satisfied for photon energies below $\approx 3\,keV$. Under these circumstances, the electric field across the atom can be assumed to be constant.

The potential $v_1(\mathbf{r}t)$ associated with a monochromatic spatially constant electric field in z-direction is given by

$$v_1(\mathbf{r}t) = E\,z\cos\omega t. \tag{81}$$

The induced density change, $\delta\rho(\mathbf{r}t)$, is most conveniently characterized by a series of multipole moments. The induced dipole moment,

$$p(t) = -\int z\delta\rho(\mathbf{r}t)d^3r, \tag{82}$$

can be expanded as[101]

$$p(t) = \alpha(\omega)\cdot\mathbf{E}\cos\omega t + \tfrac{1}{2}\beta(0) : \mathbf{EE} + \tfrac{1}{2}\beta(2\omega) : \mathbf{EE}\cos 2\omega t$$

$$+\frac{1}{4}\gamma(\omega) \vdots \mathbf{EEE}\cos\omega t+\frac{1}{4}\gamma(3\omega) \vdots \mathbf{EEE}\cos 3\,\omega t, \tag{83}$$

where the notation is meant to convey the tensorial character of the quantities α, β, γ. The first coefficient, α, is termed the *(dipole) polarizability*; β and γ denote the second-and third-order *(dipole) hyperpolarizabilities*. For ordinary laboratory fields, the electric dipole moment is adequately described by the polarizability α. In the presence of intense laser fields, hyperpolarisabilities up to γ have to be taken into account. Higher orders are negligible in present-day experiments. For spherically symmetric states β is zero.

If one considers only the linear term in Eq. (83)), the Fourier components $\rho_1(\mathbf{r}\omega)$ of the first-order density shift $\rho_1(\mathbf{r}t)$ can be treated independently and the frequency-dependent polarizability is given by

$$\alpha(\omega) = -\frac{2}{E}\int z\rho_1(\mathbf{r}\omega)d^3r. \tag{84}$$

By virtue of the Golden Rule, the photoabsorption cross section is then proportional to the imaginary part of $\alpha(\omega)$:

$$\sigma(\omega) = \frac{4\pi\omega}{c}\, Im\,\alpha(\omega). \tag{85}$$

Zangwill and Soven[35] have used the time-dependent KS scheme (51)-(52) to calculate the linear density response of rare-gas atoms to a perturbation of the form (81). For the exchange-correlation kernel f_{xc} they employed the adiabatic local density approximation (76). The resulting photoabsorption cross section (as calculated from Eqs. (84) and (85)) turns out to be in rather good agreement with experiment. As an example, we show in Fig. 3 the photoabsorption cross section of the Xe atom in the vicinity of the 4d threshold. A few general remarks are in order at this point:

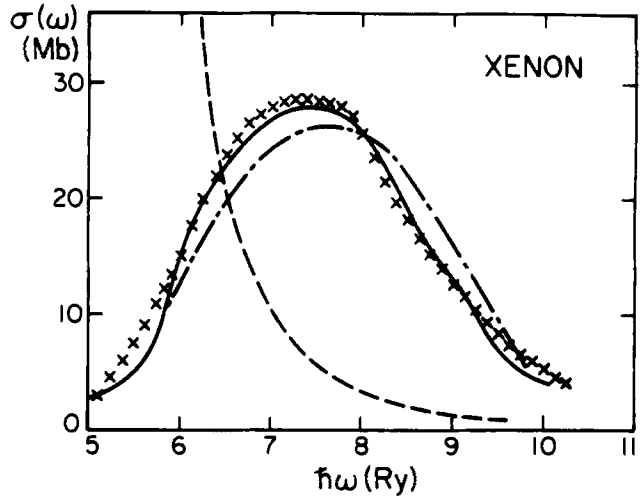


Fig. 3. Total photoabsorption cross section of the Xe atom versus photon energy in the vicinity of the 4d threshold (from Ref. 35). Solid line: selfconsistent time-dependent KS calculation; dash-dot line: self-consistent time-dependent Hartree calculation ($f_{xc} \equiv 0$); dashed line: independent particle result (Hartree and exchange-correlation kernels neglected); crosses: experimental data from Ref. 102.

i)The calculations of Zangwill and Soven do not show the characteristic autoionization resonances found in high-precision measurements. This is due to the local density approximation employed for the ground-state exchange-correlation potential in Eq. (47). Since self-Coulomb and self-exchange energies do not properly cancel in the local density approximation, the total selfconsistent potential in Eq. (47) falls off exponentially rather than like -1/r. As a consequence, the effective potential does not support the high-lying Rydberg states essential for the description of autoionization resonances. In principle, the time-dependent KS scheme (51)-(52) should be capable of describing such effects if a selfinteraction-corrected approximation for the exchange-correlation potential is used. A different approach to the

selfinteraction problem was proposed by Zangwill and Soven in a second paper,[89] where they introduced selfenergy corrections directly in the response function χ_{KS}.

ii) Within the original scheme of Zangwill and Soven, the onset of continuous absorption for each atomic subshell is given by the corresponding KS single-particle energy eigenvalue. This follows directly from the fact that the response function χ_{KS} has poles at the KS single-particle excitation energies (cf. Eq. (54)). Since the KS excitation energies do not coincide with the exact atomic excitation energies, the calculated absorption edges are typically several volts below the observed thresholds. We emphasize that this result is due to the approximation (76) used for $f_{xc}(\mathbf{r},\mathbf{r}';\omega)$. The exact function $f_{xc}(\mathbf{r},\mathbf{r}';\omega)$, formally defined as the Fourier transform of (55), has a rather intricate structure which both removes the incorrect poles of χ_{KS} and gives rise to poles at the correct quasi-particle excitation energies.

iii) It is of interest to estimate the overall importance of the exchange-correlation kernel f_{xc}. The dash-dot curve in Fig. 3 shows the result obtained with f_{xc} set equal to zero. Deviations from the full calculation (solid curve) are about 10%. Clearly the full calculation agrees much better with the experimental values.

iv) Zangwill and Soven have employed the adiabatic local density approximation (76), in which $f_{xc}(\mathbf{r},\mathbf{r}';\omega)$ is replaced by $f_{xc}(\mathbf{r},\mathbf{r}';\omega = 0)$. On the basis of the parametrization (78)-(79) of Gross and Kohn,[83] the importance of the frequency dependence of the exchange-correlation kernel has been estimated: In the frequency range relevant for atomic photoabsorption, the ω-dependent function f_{xc} is expected[75] to differ by at most 2-6% from its zero-frequency value employed in the calculation of Zangwill and Soven. Since the complete neglect of f_{xc} leads to deviations of only 10%, the effect of the frequency-dependence of f_{xc} can be estimated to be smaller than 1% in the photoabsorption cross

section of atoms.

The scheme of Zangwill and Soven has also been applied to calculations of the photoabsorption of small molecules.[90,91] In particular, a previously unexplained double-peak structure near $15\,eV$ in the photoemission cross section of acetylene could be described successfully.

The adiabatic exchange-correlation functional (74) has also been employed in calculations of the third-order hyperpolarizabilities defined in Eq. (83). The results obtained by Zangwill[103] for rare-gas atoms are in good agreement with experiment. According to Senatore and Subbaswamy,[104] however, a term was missing in Zangwill's calculation. When they include this term, the agreement with experiment is reduced considerably: The calculated third-order hyperpolarizabilities of rare-gas atoms are too large typically by a factor of 2.

The response problem of surface electrons is to find the currents and charge densities induced by perturbing electromagnetic fields, and to calculate the resulting modifications of these fields.

The Hamiltonian governing the electronic dynamics is

$$\hat{H} \equiv \sum_i \tfrac{1}{2}(\mathbf{p}_i - \tfrac{1}{c}\mathbf{A}_1(\mathbf{r}_i,t))^2 + \sum_i (v_0(\mathbf{r}_i) + v_1(\mathbf{r}_i,t)) + \tfrac{1}{2}\sum_{i\neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{86}$$

where $v_0(\mathbf{r})$ is the potential due to the ion cores and $v_1(\mathbf{r},t)$, $\mathbf{A}_1(\mathbf{r},t)$ are the perturbing potentials in the transverse gauge ($\nabla\cdot\mathbf{A}_1 = 0$). The primary objective is to find the first-order density and current changes, $\rho_1(\mathbf{r},t)$ and $\mathbf{j}_1(\mathbf{r},t)$, caused by v_1 and $\mathbf{A}_1$.

In most theoretical work, as well as in this section, the ionic potential, $v_0(\mathbf{r})$, is replaced by the potential due to a uniform positive charge background in a half-space, say $z > 0$. This is the so-called jellium model for metal surfaces. In this model there are two intrinsic microscopic length scales, the inverse Fermi wave-number, k_F^{-1}, and the Thomas-Fermi screening length ($\approx$ surface thickness), κ_{TF}^{-1}. In the

present context, both lengths are typically of the order $a \approx 10^{-8}$ *cm*.

In the most important applications the perturbing potentials v_1 and $\mathbf{A}_1$ vary on a length scale, ℓ, which satisfies $\ell \gg a$. Examples are the scalar potential $v_1(\mathbf{r})$ due to an external charge at a distance $z \gg a$, or the vector potential, $\mathbf{A}_1(\mathbf{r},t)$, associated with a light wave of wavelength $\lambda \gg a$. The corresponding n_1 and j_1 vary on the scale of ℓ in the x-y plane but, because of the abrupt drop of the unperturbed density at the surface (on the scale of a), they vary on the short scale a in the z-direction. Formal arguments due to Feibelman[105] have shown that, to leading order in a/ℓ, the effect of the surface on the electromagnetic fields far from the surface ($|z| \gg a$) is entirely characterized by two complex frequency-dependent lengths, $d_{\|}(\omega)$ and $d_{\perp}(\omega)$.

Actually, for the jellium model $d_{\|}(\omega) \equiv 0$.* To define $d_{\perp}(\omega)$ we Fourier analyze all physical quantities parallel to the surface, in the x-y plane. For example, a Fourier component of the induced charge density becomes

$$\rho_1(\mathbf{r},\omega) \equiv \rho_1(z,\omega)e^{i\mathbf{q}_{\|}\cdot\mathbf{r}}, \tag{87}$$

where $\mathbf{q}_{\|} = (q_x, q_y, 0)$ (and $|\mathbf{q}_{\|}|\cdot a \ll 1$). Then $d_{\perp}(\omega)$ is given by

$$d_{\perp}(\omega) \equiv \int z\rho_1(z,\omega)dz / \int \rho_1(z,\omega)dz, \tag{88}$$

i.e. it is the (complex) center of mass of the induced surface charge. $d_{\perp}(\omega)$ is the generalization of the static image plane introduced by Lang and Kohn.[106]

This $d_{\perp}(\omega)$ is then the object of quantitative calculations. They

* This result has been obtained in the random phase approximation in Ref. 105. It is easily established as a rigorous many-body result for the jellium model (W. Kohn, unpublished).

require the density response, $\rho_1(z,\omega)$, to a uniform external electric field perpendicular to the surface. The calculation was first carried out in the random phase approximation (RPA), equivalent to time-dependent Hartree theory, in which effects of exchange and correlation are neglected in the self consistent perturbing potential (52). These calculations led to very interesting results not present in classical Maxwell theory, the surface photo effect, excitations of bulk and surface plasmons, and hints of an additional surface-localized plasmon. They help to explain experimental data on the photo-yield spectrum of Al.[107]

Some recent calculations[92–98] have employed the time-dependent response theory of Gross and Kohn, including exchange-correlation effects either within the adiabatic approximation (75) or the frequency-dependent parametrization (78)-(79). They have confirmed the qualitative conclusions of the RPA calculations, with quantitative modifications generally of the order of 10%.

Finally we mention some recent formal work on the response properties of real crystal surfaces (not the jellium model).[108] Here again the response properties of the surface are characterized by complex d-parameters having the dimension of a length. Quantitative calculations have been reported[108] to be in progress.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. DMR87-03434 and by the U. S. Office of Naval Research under Contract No. N00014-89-J-1530. One of us (E.K.U.G.) acknowledges a Heisenberg fellowship of the Deutsche Forschungsgemeinschaft.

REFERENCES

1. P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
2. W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
3. L. H. Thomas, Proc. Cambridge Phil. Soc. **23**, 542 (1927).
4. E. Fermi, Rend. Accad. Naz. Linzei **6**, 602 (1927).
5. D. R. Hartree, Proc. Cambridge Phil. Soc. **24**, 89 (1928).
6. J. C. Slater, Phys. Rev. **81**, 385 (1951).
7. E. Runge and E.K.U. Gross, Phys. Rev. Lett. **52**, 997 (1984).
8. F. Bloch, Z. Phys. **81**, 363 (1933).
9. M. Horbatsch and R. M. Dreizler, Z. Phys. A **300**, 119 (1981).
10. M. Horbatsch and R. M. Dreizler, Z. Phys. **A308**, 329 (1982).
11. B. M. Deb and P. K. Chattaraj, Phys. Rev. **A39**, 1696 (1989).
12. G. Holzwarth, Phys. Lett. **66B**, 29 (1977).
13. G. Holzwarth, Phys. Rev. **C16**, 885 (1977).
14. H. Sagawa and G. Holzwarth, Prog. Theor. Phys. **59**, 1213 (1978).
15. G. Holzwarth, in *Density Functional Methods in Physics*, edited by R. M. Dreizler and J. da Providencia, p. 381 (Plenum Press, New York, 1985).
16. J. P. da Providencia and G. Holzwarth, Nucl. Phys. **A439**, 477 (1985).
17. J. A. Ball, J. A. Wheeler, and E. L. Fireman, Rev. Mod. Phys. **45**, 333 (1973).
18. P. Malzacher and R. M. Dreizler, Z. Phys. **A307**, 211 (1982).
19. P. Zimmerer, N. Grün, and W. Scheid, Phys. Lett. **A134**, 57 (1988).
20. W. Brandt and S. Lundqvist, Phys. Rev. **139**, A612 (1965).
21. A. Sen and E. G. Harris, Phys. Rev. **A3**, 1815 (1971).
22. N. H. March and M. P. Tosi, Proc. Roy. Soc. London **A330**, 373 (1972).
23. N. H. March and M. P. Tosi, Phil. Mag. **28**, 91 (1973).
24. N. H. March, in *Elementary Excitations in Solids, Molecules, and Atoms,* edited by J. T. Devreese, A. B. Kunz, and T. C. Collins, p. 97 (Plenum Press, New York, 1974).
25. S. C. Ying, Nuovo Cim. **23B**, 270 (1974).
26. P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 376 (1930).
27. *Time-Dependent Hartree Fock Method*, edited by P. Bonche and P. Quentin (Editions de Physique, Paris, 1979).
28. J. W. Negele, Rev. Mod. Phys. **54**, 913 (1982).

29. K. C. Kulander, K. R. Sandhya Devi, and S. E. Koonin, Phys. Rev. **A25**, 2968 (1982).
30. W. Stich, H.-J. Lüdde, and R. M. Dreizler, Phys. Lett. **99A**, 41 (1983).
31. K. R. Sandhya Devi and J. D. Garcia, Phys. Rev. **A30**, 600 (1984).
32. W. Stich, H.-J. Lüdde, and R. M. Dreizler, J. Phys. **B18**, 1195 (1985).
33. D. Tiszauer and K. C. Kulander, Phys. Rev. **A29**, 2909 (1984).
34. V. Peuckert, J. Phys. **C11**, 4945 (1978).
35. A. Zangwill and P. Soven, Phys. Rev. **A21**, 1561 (1980).
36. B. M. Deb and S. K. Ghosh, J. Chem. Phys. **77**, 342 (1982).
37. S. K. Ghosh and B. M. Deb, Chem. Phys. **71**, 295 (1982).
38. S. K. Ghosh and B. M. Deb, Theor. Chim. Acta **62**, 209 (1983).
39. S. K. Ghosh and B. M. Deb, J. Mol. Struct. **103**, 163 (1983).
40. L. J. Bartolotti, Phys. Rev. **A24**, 1661 (1981).
41. L. J. Bartolotti, Phys. Rev. **A26**, 2243 (1982).
42. L. J. Bartolotti, J. Chem. Phys. **80**, 5687 (1984).
43. L. J. Bartolotti, Phys. Rev. **A36**, 4492 (1987).
44. L. F. Errea, L. Méndez, A. Riera, M. Yáñez, J. Hanssen, C. Harel, and A. Salin, J. Physique **46**, 719 (1985).
45. I. L. Cooper, A. S. Dickinson, S. K. Sur, and C. T. Ta, J. Phys. **B20**, 2005 (1987).
46. A. Henne, H.-J. Lüdde, A. Toepfer, and R. M. Dreizler, Phys. Lett. **A124**, 508 (1987).
47. W. Fritsch and C. D. Lin, Phys. Lett. **A123**, 128 (1987).
48. J. F. Reading and A. L. Ford, Phys. Rev. Lett. **58**, 543 (1987).
49. J. F. Reading and A. L. Ford, J. Phys. **B20**, 3747 (1987).
50. A. Toepfer, B. Jacob, H.-J. Lüdde, and R. M. Dreizler, Phys. Lett. **93A**, 18 (1982).
51. E.K.U. Gross and R. M. Dreizler, in *Density Functional Methods in Physics*, edited by R. M. Dreizler and J. da Providencia, p. 81 (Plenum Press, New York, 1985).
52. A. Toepfer, H.-J. Lüdde, B. Jacob, and R. M. Dreizler, J. Phys. B **18**, 1969 (1985).
53. A. Toepfer, A. Henne, H.-J. Lüdde, M. Horbatsch, and R. M. Dreizler, Phys. Lett. A **126**, 11 (1987).
54. P. M. Koch, K. A. H. van Leeuwen, O. Rath, D. Richards, and R. V. Jensen, in *The Physics of Phase Space*, edited by Y. S. Kim and W. W. Zachary, Lecture Notes in Physics **278**, p. 106, (Springer-Verlag, Berlin, 1987).

55. G. Casati and I. Guarneri, in *Quantum Chaos and Statistical Nuclear Physics*, edited by T. H. Seligman and H. Nishioka, Lecture Notes in Physics **263**, p. 238 (Springer-Verlag, Berlin, 1986).
56. K. C. Kulander, Phys. Rev. A **35**, 445 (1987).
57. K. C. Kulander, Phys. Rev. A **36**, 2726 (1987).
58. E. H. Lieb, in *Density Functional Methods in Physics*, edited by R. M. Dreizler and J. da Providencia, p. 31 (Plenum Press, New York, 1985).
59. B.-X. Xu and A. K. Rajagopal, Phys. Rev. A **31**, 2682 (1985).
60. A. K. Dhara and S. K. Ghosh, Phys. Rev. A **35**, 442 (1987).
61. M. Levy, Phys. Rev. **A26**, 1200 (1982).
62. M. Levy, Proc. Natl. Acad. Sci. USA **76**, 6062 (1979).
63. H. Kohl and R. M. Dreizler, Phys. Rev. Lett. **56**, 1993 (1986).
64. S. K. Ghosh and A. K. Dhara, Phys. Rev. **A38**, 1149 (1988).
65. K. L. Liu and S. H. Vosko, Can. J. Phys., in press (1989).
66. T.-C. Li and P.-Q. Tong, Phys. Rev. **A34**, 529 (1986).
67. T.-C. Li and P.-Q. Tong, Phys. Rev. **A31**, 1950 (1985).
68. T.-C. Li and Y. Li, Phys. Rev. **A31**, 3970 (1985).
69. T. K. Ng, Phys. Rev. Lett. **62**, 2417 (1989).
70. W. Yang, Phys. Rev. **A38**, 5512 (1988).
71. An existence proof referring to the time-dependent linear density responses of a somewhat different class of time-dependent perturbations was given by T. K. Ng and K. S. Singwi, Phys. Rev. Lett. **59**, 2627 (1987).
72. D. Mearns and W. Kohn, Phys. Rev. **A35**, 4796 (1987).
73. E.K.U. Gross, D. Mearns, and L. N. Oliveira, Phys. Rev. Lett. **61**, 1518 (1988).
74. S. Ichimaru, Rev. Mod. Phys. **54**, 1017 (1982).
75. N. Iwamoto and E.K.U. Gross, Phys. Rev. **B35**, 3003 (1987).
76. D. M. Ceperley, Phys. Rev. **B18**, 3126 (1978).
77. S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
78. R. W. Shaw, J. Phys. **C3**, 1140 (1970).
79. G. Niklasson, Phys. Rev. **B10**, 3052 (1974).
80. A. A. Kugler, J. Stat. Phys. **12**, 35 (1975).
81. A. Holas and K. S. Singwi, Phys. Rev. **B40**, 158 (1989).
82. A. J. Glick and W. F. Long, Phys. Rev. **B4**, 3455 (1971).
83. E.K.U. Gross and W. Kohn, Phys. Rev. Lett. **55**, 2850 (1985); Erratum: ibid. **57**, 923 (1986).
84. N. Iwamoto, Phys. Rev. **A30**, 2597 (1984).
85. N. Iwamoto, Phys. Rev. **A30**, 3289 (1984).

86. P. F. Byrd and M. D. Friedman, *Handbook of Eliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).
87. B. Dabrowski, Phys. Rev. B**34**, 4989 (1986).
88. A. Zangwill and P. Soven, Phys. Rev. Lett. **45**, 204 (1980).
89. A. Zangwill and P. Soven, Phys. Rev. B**24**, 4121 (1981).
90. Z. H. Levine and P. Soven, Phys. Rev. Lett. **50**, 2074 (1983).
91. Z. H. Levine and P. Soven, Phys. Rev. A**29**, 625 (1984).
92. A. Liebsch, Phys. Rev. B**36**, 7378 (1987).
93. J. F. Dobson and G. H. Harris, J. Phys. C**20**, 6127 (1987).
94. J. F. Dobson and G. H. Harris, J. Phys. C**21**, L729 (1988).
95. P. Gies and R. R. Gerhardts, Phys. Rev. B**36**, 4422 (1987).
96. P. Gies and R. R. Gerhardts, J. Vac. Sci. Technol. A**5**, 936 (1987).
97. P. Gies and R. R. Gerhardts, Phys. Rev. B**37**, 10020 (1988).
98. K. Kempa and W. L. Schaich, Phys. Rev. B**37**, 6711 (1988).
99. T. Ando, Z. Phys. B**26**, 263 (1977).
100. Z. H. Levine and D. C. Allan, Phys. Rev. Lett. **63**, 1719 (1989).
101. P. W. Langhoff, S. T. Epstein, and M. Karplus, Rev. Mod. Phys. **44**, 602 (1972).
102. R. Haensel, G. Keitel, P. Schreiber, and C. Kunz, Phys. Rev. **188**, 1375 (1969).
103. A. Zangwill, J. Chem. Phys. **78**, 5926 (1983).
104. G. Senatore and K. R. Subbaswamy, Phys. Rev. A **35**, 2440 (1987).
105. P. J. Feibelman, Prog. Surf. Sci. **12**, 287 (1982).
106. N. D. Lang and W. Kohn, Phys. Rev. B**7**, 3541 (1973).
107. H. J. Levinson, E. W. Plummer and P. J. Feibelman, Phys. Rev. Lett. **43**, 952 (1979).
108. W. L. Schaich and Wei Chen, Phys. Rev. B**39**, 10714 (1989).

INTEGRAL FORMULATION OF DENSITY-FUNCTIONAL THEORY

Weitao Yang

Department of Chemistry
University of California
Berkeley, CA 94720, U.S.A.

ABSTRACT

The Hohenberg-Kohn-Sham density-functional theory is reformulated in terms of explicit relations between the electron density and the effective potential through the use of Feynman path integrals. In this formulation electron density is the only basic variable as in the Thomas-Fermi theory and orbitals are not needed. Possible applications to calculations in large molecules and the present limitations of the method are discussed.

1. ELECTRON DENSITY AS THE BASIC VARIABLE

In their seminal paper [1] Hohenberg and Kohn proved that electron density alone can be used as the basic variable for electronic structure problems. Their work not only provided the theoretical foundation for the Thomas-Fermi theory and its extensions [2,3], but also led to the Kohn-Sham formulation [4], and to the great success of the density-functional theory [5-12].

The Kohn-Sham formulation has been extensively applied to molecular and solid systems [11,12]. In essence, the Kohn-Sham method uses a set of auxiliary orbitals to determine the noninteracting electron kinetic energy (which is the dominant part of the exact kinetic energy), and includes the rest of the energy functional in the Kohn-Sham effective potential. The resultant working equations are a set of coupled single electron equations, which are normally turned into algebraic eigenvalue

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

problems through the use of basis functions [11,12]. Within the Kohn-Sham method, electron density is parameterized in terms of orbitals and not used directly as the basic variable. The success of this method lies in the accuracy associated with treating the dominant part of kinetic energy correctly. However, the costs for this gain in accuracy are that one is forced to deal with orbitals individually, and hence to have a difficulty similar to the Hartree-Fock method when it comes to calculations for large molecules.

The alternative method is to have the electron density directly as the basic variable, originating from the Thomas-Fermi model [2,3,6]. In the Thomas-Fermi method, the noninteracting electron kinetic enery is approximated explicitly as a functional of the electron density, leading to an equation for the electron density. The idea is indeed wonderful: one needs just a single equation for large systems with arbitrary numbers of electrons. Unfortunately, the approximate kinetic energy functional is still not accurate enough for this approach to have much value in practical applications, despite countless efforts. Inclusion of gradients of electron density in the functional can make partial correction, as demonstrated in recent studies [13-19]. There are other direct approaches which use the full Weizsacker kinetic energy as the main ingredient in place of the Thomas-Fermi kinetic energy [20-28]; the result is an equation for the square root of the electron density [22,24-28]. There is also a different approach to the problem by constructing an approximate equation for the local effective potential [29]. However, the accurate approximation of kinetic energy as an explicit functional of the electron density remains to be one of the most intriguing and challenging problems in density-functional theory.

Significant progress has been made recently in the direct approach, not on the kinetic energy functional itself, but on the explicit functional expression of electron density in terms of the Kohn-Sham effective potential [30-34]. This is the integral formulation of the Kohn-Sham theory, which is the focus of the present paper.

The method to bypass orbitals has been known for a long time: in a formal way via the use of the single-particle Green's function [35]. Golden suggested the use of operator product approximation for improving the Thomas-Fermi model [36]. The product approximation for the Green's function was shown to give convergent and correct result for noninteracting electrons in a linear potential by Handler and Wang [37] and in a quadratic potential by Handler [30]. Handler also gave an expression for the density matrix in terms of the local potential in the product representation for the Green's function via the inverse Laplace transform [30]. Starting with the path integral representation of the Green's function, which is essentially equivalent to the operator product representation,

Harris and Pratt carried out the inverse Laplace transform analytically and expressed the electron density as a multidimensional integral within the framework of the many-electron Hartree theory [31]. They also proposed the use of such a relation as an integral equation for the local Hartree potential. Building on these studies, Yang established the relation between electron density and local potential in the Kohn-Sham theory, thereby generalizing the work of Harris and Pratt to include the effects of electron exchange and correlation [32]. The possibility of this generalization was also mentioned by Pratt, Hoffman and Harris [33]. Yang also gave the spin-polarized version of the theory [32], and constructed several forms of the functional relation between density and potential based on different path integral representations [34].

Unlike the gradient correction for the Thomas-Fermi theory, which is a perturbative expansion [38-39], the integral formulation of the Kohn-Sham theory is based on convergent sequences of approximation for the Green's function from the Feynman path integral approach [40,41]. Therefore, it uses electron density as the basic variable in an explicit way with accuracy that can be controlled and improved by the convergent path integral representation of the Green's function.

2. THE INTEGRAL FORMULATION

According to Hohenberg-Kohn-Sham density functional theory, the total electronic energy of a N-electron system is given by the minimum of the functional

$$E[\rho'] = T_s[\rho'] + \frac{e^2}{2}\int \frac{\rho'(\vec{r})\rho'(\vec{r}')}{|\vec{r}-\vec{r}'|}\, d\vec{r}d\vec{r}' + \int v(\vec{r})\rho'(\vec{r})d\vec{r} + E_{xc}[\rho'] \qquad (1)$$

subject to the constraint of density normalization

$$\int \rho'(\vec{r})d\vec{r} = N, \qquad (2)$$

where $T_s[\rho']$ is the Kohn-Sham noninteracting kinetic energy, $v(\vec{r})$ is the external potential (the nuclear potential in molecules), and $E_{xc}[\rho']$ is the exchange-correlation energy. The Euler equation for the ground state density $\rho(\vec{r})$ is

$$\frac{\delta T_s[\rho]}{\delta \rho(\vec{r})} + v_{eff}(\vec{r}) = \mu, \tag{3}$$

where μ is the Lagrange multiplier for the normalization constraint and v_{eff} is the Kohn-Sham local potential

$$v_{eff}(\vec{r}) = v(\vec{r}) + e^2 \int \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} d\vec{r}' + \frac{\delta E_{xc}[\rho]}{\delta \rho(\vec{r})} . \tag{4}$$

In principle, Eq. (3) is all that one needs to solve to obtain the ground state density and energy. However, the exact forms of $T_s[\rho]$ and $E_{xc}[\rho]$ still evade us. Kohn and Sham [4] solved the problem of $T_s[\rho]$ indirectly, but exactly by indroducing a set of single-electron orbitals satisfying the equation

$$\hat{H}\psi_i(\vec{r}) = [- \frac{\hbar^2}{2m} \nabla^2 + v_{eff}(\vec{r})] \psi_i = \varepsilon_i \psi_i(\vec{r}), \tag{5}$$

and giving the electron density

$$\rho(\vec{r}) = 2 \sum_{i=1}^{N/2} |\psi_i(\vec{r})|^2, \tag{6}$$

where the sum is over the N/2 lowest eigenstates of $\hat{H}$. Closed-shell structure is assumed here, but the generalization to spin-polarized situation is straightforward.

The Kohn-Sham method replaces Eq. (3), a single equation for density, with Eq. (5), eigenvalue problem for N/2 lowest eigenstates. This is a trade of simplicity for accuracy. However, it is the use of these N/2 orbitals that limits the application of the Kohn-Sham method to molecules with number of electrons N of the the order 10^2. In contrast to the Kohn-Sham method, direct methods can be regarded as techniques which compute (approximately) the total electron density without using the N/2 orbitals as expressed in Eq. (6). The motivation for such an approach is to bypass the computational effort of determining the N/2 orbitals, which is of the order N^3 in CPU times. The Thomas-Fermi theory and its extensions are direct approaches to Eq. (3), which approximate $T_s[\rho]$. The integral formulation is also a direct approach, but starting from the Kohn-Sham orbitals, as described below.

The ground state electron density of Eq. (6) can be expressed as the diagonal of the corresponding density matrix

$$\rho(\vec{r}, \vec{r}') = 2 \sum_{i=1}^{N/2} \psi_i(\vec{r})\psi_i^*(\vec{r}') = 2 \langle \vec{r} | \eta(\varepsilon_F - \hat{H}) | \vec{r}' \rangle, \quad (7)$$

where ε_F is the Fermi energy, the value of which lies between the highest occupied and the lowest unoccupied orbital energies, and $\eta(x)$ is the unit step function. A well-known formula relates this density matrix to the single-particle Green's function [35]; namely,

$$\rho(\vec{r}, \vec{r}') = \frac{1}{\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} d\beta \frac{1}{\beta} \exp(\beta\varepsilon_F) G(\vec{r}, \vec{r}'; \beta) \quad (8)$$

where the Green's function is

$$G(\vec{r}, \vec{r}'; \beta) = \langle \vec{r} | \exp(- \beta\hat{H}) | \vec{r}' \rangle, \quad (9)$$

$\hat{H}$ is the effective single-particle Hamitonian of Eq. (5), and γ is any positive constant. The integration on the complex-β plane in Eq. (8) is the inverse Laplace transform. Eq. (8) has been used in conjunction with a $\hbar$-semiclassical expansion for the Green's function to obtain the gradient expansion for the kinetic energy functional [38-39]. The Feynman path integral formulation for the Green's function is used here [40,41]; namely,

$$G(\vec{r}, \vec{r}'; \beta) = \int \mathcal{D}\vec{x}(\tau) \exp(- S[\vec{x}(\tau)]), \quad (10)$$

where the classical action is

$$S[\vec{x}(\tau)] = \beta \int_0^1 \{ \frac{m}{2\hbar^2\beta^2} [\dot{\vec{x}}(\tau)]^2 + v_{eff}(\vec{x}(\tau)) \} d\tau , \quad (11)$$

and the path integration is over all possible paths satisfying the boundary conditions

$$\vec{x}(0) = \vec{r}', \quad \vec{x}(1) = \vec{r} . \quad (12)$$

In the polygonal, or discrete path representation, the electron paths are represented as piecewise straight lines; this has been used to express the electron density in terms of the

local potential, in simple model [30], in many electron Hartree theory [31] and in the Kohn-Sham theory [32]. There are other ways of carrying out the path integrations [40,41], leading to corresponding representations of density in terms of potential [34]. Some may have better convergence property. The use of Fourier path representation is described here [34]; it also has been used in quantum dynamics and statistical mechanics (for example, [42-44]).

In the Fourier path representation, all paths with boundary condition Eq. (12) are expanded in terms of Fourier series [40,42-44]

$$\vec{x}_p(\tau) = \vec{r}' + (\vec{r} - \vec{r}')\tau + \frac{\sqrt{2}}{\pi} \sum_{k=1}^{p-1} \vec{z}_k \frac{\sin(k\pi\tau)}{k} . \qquad (13)$$

Integration over all such paths is carried out by integration over all possible real values of the Fourier coefficients z_k. The kinetic energy part of the classical action can be written explicitly in terms of z_k and the final form of the Green's function is

$$G_p(\vec{r},\vec{r}';\beta) = \left(\frac{m}{2\pi\hbar^2\beta^2}\right)^{3p/2} \int d\vec{z}_1 \cdots d\vec{z}_{p-1} \exp\{-\frac{m}{2\hbar^2\beta}[\sum_{k=1}^{p-1} |\vec{z}_k|^2 + (\vec{r} - \vec{r}')^2] - \beta\int_0^1 v_{eff}(\vec{x}_p(\tau))\, d\tau\}. \qquad (14)$$

This expression is an approximation to the exact one, Eq. (10), because of the truncation of the Fourier series in Eq. (13). The accuracy of Eq. (14) increases with increasing value of p. In other words, Eq. (14) is known to be convergent with respect to p [40,42-44].

Now, the Fourier path integral of Eq. (14) is inserted into Eq. (8), leading to the corresponding density matrix

$$\rho_p(\vec{r},\vec{r}') = 2\int d\vec{z}_1 \cdots d\vec{z}_{p-1} \left[\frac{k_p}{2\pi\ell_p}\right]^{3p/2} J_{3p/2}(k_p\ell_p)\eta(k_p^2), \qquad (15)$$

where

$$\ell_p^2 = |\vec{r} - \vec{r}'|^2 + \sum_{k=1}^{p-1} |\vec{z}_k|^2 , \qquad (16)$$

$$k_p^2 = \frac{2m}{\hbar^2} [\, \varepsilon_F - \int_0^1 v_{eff}(\vec{x}_p(\tau))\, d\tau]. \qquad (17)$$

The diagonal part of Eq. (15) gives the electron density in

terms of the effective potential

$$\rho_p(\vec{r}) = \rho_p(\vec{r}, \vec{r}) = \rho_p[\varepsilon_F - v_{eff}; \vec{r}] \ . \qquad (18)$$

Eq. (15) is the central result of the integral formulation of the Hohenberg-Kohn-Sham theory within the Fourier path integral representation. Eq. (18) can be considered as solution to the Euler equation (3) at p-level approximation. Since the effective potential itself depends on the density, Eq. (18) has to be solved self-consistently, with the Fermi energy ε_F determined by the normalization constraint. Although the corresponding functional form of $T_s[\rho]$ is still unknown, the kinetic energy can be obtained by the used of Eq. (15) with v_{eff} given by the self-consistent density. There are several ways of doing this [34].

At the lowest level of approximation, p=1, the electron path of Eq. (13) is a straight line and there is no integration in Eq. (15). Correspondingly, Eq. (18) becomes the local Thomas-Fermi relation. In other words, Thomas-Fermi theory neglects any detail of the electron dynamics.

3. SIGNIFICANCE AND PRESENT LIMITATIONS

The integral formulation provides successive, in principle convergent approximations beyond the Thomas-Fermi theory, in contrast to the conventional gradient expansion which is divergent. It uses the electron density as the basic variable in the same way as the Thomas-Fermi theory. Calculations for noninteracting fermions in a one-dimensional harmonic potential indicate fast convergence: quantum shell structure, which is absent from the Thomas-Fermi theory, appear in the second-level approximation and further improvement is achieved with the third-level approximation [45]. For a Monte Carlo calculation on a harmonic potential system see Ref 46.

With the formal and explicit relation of density and local Kohn-sham effective potential established, one certainly would like to explore its practical applications, in particular to large systems where the solution of the Kohn-Sham equation is prohibitive. This expectation arises from the fact that the number of electrons only plays the role of normalization constant for the density in the integral formulation.

When it comes to numerical implementation, however, Eq. (15) present difficulty that has not been overcome so far; namely, it is a multidimensional integral with oscillatory integrand. If one uses numerical quadrature methods,

calculations of electron density can be done for the approximation at p=2, which should improve the Thomas-Fermi model, but still can be of poor accuracy. The direct numerical integrations would be very difficult (in terms of the computation time requirement) to carry out for approximations beyond p=2. At p=3, there are nine-dimensional integrals involved. Another possible direction would be to use some type of Monte Carlo methods. But the oscillatory integrand in Eq. (15) makes it difficult to apply Monte Carlo methods, which normally requires a positive integrand for it to be efficient. Problems of this nature also occur in path integral approach to quantum dynamics, where one needs to evaluate Eq. (14) for complex values of β; there are some initial progress using Monte Carlo method [47,48].

The integral formulation can also be considered as a method of bypassing the Hamiltonian matrix construction and diagonization associated with the conventional method of solving eigenvalue problems in Kohn-Sham theory. This idea has also been extended to the finite-temperature Kohn-Sham theory [49] and to linear response problems [50].

The advance of the method now depends on the development of efficient methods for doing the multidimensional integration involved in Eq. (15). It is another challenging problem.

ACKNOWLEDGEMENTS

The author would like to thank Professor William Miller for support. Research was funded by the National Science Fundation, Grant No. CHE 84-16345.

REFERENCES

1. P. Hohenberg and W. Kohn, Phys. Rev. 136, B 864 (1964).
2. L.H. Thomas, Proc. Camb. Phil Soc. 23, 542 (1926).
3. E. Fermi, Z. Phys. 48, 73 (1928).
4. W. Kohn and L.J. Sham, Phys. Rev. A 140, 1133 (1965).
5. W. Kohn and P. Vashista, in Theory of the Inhomogeneous Electron Gas, edited by S. Lundqvist and N.H. March (Plenum, New York, 1983), p 79.
6. N.H. March, in Theory of the Inhomogeneous Electron Gas, edited by S. Lundqvist and N.H. March (Plenum, New York, 1983), p 1.
7. A.K. Rajagopal, Adv. in Chem. Phys. 41, 59 (1980).
8. R.G. Parr, Ann. Rev. Phys. Chem. 34, 631 (1983).

9. M. Levy and J.P. Perdew, in Density Functional Methods in Physics, edited by R.H. Dreizler and J. da Providencia (Plenum, New York, 1983), p 11.
10. B.M. Deb and S.K. Ghosh, in The Single-Particle Density in Physics and Chemistry, edited by N.H. March and B.M. Deb, (Academic Press, London, 1987).
11. A.R. Williams and U. von Barth in Theory of the Inhomogeneous Electron Gas, edited by S. Lundquist and N.H. March (Plenum, New York, 1983), p 189.
12. R.O. Jones, in Ab initio Methods in Quantum Chemistry, Part 1, edited by K.P. Lawley (Wiley, New York, 1987), p 413.
13. N.L. Allan, C.G. West, D.L. Cooper, P.J. Grout and N.H. March, J. Chem. Phys. 83, 4562 (1985).
14. C. Lee and S.K. Ghosh, Phys. Rev. A 33, 3506 (1986).
15. M.L. Plumer and M.J. Stott, J. Phys. C 18, 4143 (1985).
16. (a) W. Yang, Phys. Rev. A 34, 4575 (1986). (b) W. Yang, R.G. Parr and C. Lee, Phys. Rev. A 34, 4586 (1986).
17. A.E. Depristo and J.D. Kress, Phys. Rev. A 35, 442 (1987).
18. J.P. Perdew, M. Levy, G.S. Painter, S. Wei and J. Lagowski, Phys. Rev. B37, 838 (1988).
19. B.M. Deb and P.K. Chattaraj, Phys. Rev. A 37, 4030 (1988).
20. P.K. Acharya, L.J. Bartolotti, S.B. Sears and R.G. Parr, Proc. Natl. Acad. Sci. USA 77, 6978 (1980).
21. J.L. Gazquez and J. Robles, J. Chem. Phys. 76, 1467 (1982).
22. B.M. Deb and S.K. Ghosh, Int. J. Quantum Chem. 23, 1 (1983).
23. (a) E.V. Ludena, J. Chem. Phys. 79, 6174 (1983); (b) M.L. Zorita, J.A. Alonso and L.C. Balbas, Int. J. Quantum Chem. 28, 393 (1985).
24. M. Levy, J.P. Perdew and V. Sahni, Phys. Rev. A 30, 2745 (1984).
25. N.H. March, Phys. Lett. A 113, 476 (1926).
26. C. Herring and M. Chopra, Phys. Rev. A 37, 31 (1988).
27. M. Levy and H. Ou-Yang, Phys. Rev. A 38, 625 (1988).
28. J. Cioslowski, J. Chem. Phys. 89, 4871 (1988).
29. B.-G. Englert and J. Schwinger, (a) Phys. Rev. A 29, 2331 (1984); (b) Phys. Rev. A 29, 2339 (1984); (c) Phys. Rev. A 29, 2353 (1984).
30. G. Handler, J. Chem. Phys. 58, 1 (1973).
31. R.A. Harris and L.R. Pratt, 82, 856 (1985).
32. W. Yang, Phys. Rev. Lett. 59, 1569 (1987).
33. L.R. Pratt, G.G. Hoffman and R.A. Harris, J. Chem. Phys. 88, 1818 (1988).
34. W. Yang, Phys. Rev. A 38, 5494 (1988).
35. S. Golden, Rev. Mod. Phys. 32, 322 (1960).
36. S. Golden, Phys. Rev. 105, 604 (1957).
37. G. Handler and P.S.C. Wang, J. Chem. Phys. 56, 1546 (1972)
38. C.H. Hodges, Can. J. Phys. 51, 1428 (1973).
39. D.R. Murphy, Phys. Rev. A 24, 1682 (1981).

40. R.P. Feynman and A.R. Hibbs, Quantum Mechanics and Path Integrals (McGraw-Hill, New York, 1965).
41. L.S. Schulman, Techniques and Applications of Path Integration (Wiley-Interscience, New York, 1981).
42. W.H. Miller, J. Chem. Phys. 63, 1166 (1975).
43. D. Freeman and J.D. Doll, J. Chem. Phys. 80, 5709 (1984).
44. R.D. Coalson, J. Chem. Phys. 85, 926 (1986).
45. J. Cioslowski and W. Yang, unpublished calculations.
46. G.G. Hoffman, L.R. Pratt and R.A. Harris, Chem. Phys. Lett. 148, 313 (1988).
47. N. Makri and W.H. Miller, Chem. Phys. Lett., 139, 10 (1987).
48. J.D. Doll, D.L. Freeman and M.J. Gillan, Chem. Phys. Lett., 143, 277 (1988).
49. W. Yang, Phys. Rev. A 38, 5504 (1988).
50. W. Yang, Phys. Rev. A 38, 5512 (1988).

DENSITY FUNCTIONAL REFORMULATION OF MOLECULAR ORBITAL THEORIES

Jerzy Cioslowski

Los Alamos National Laboratory, Theoretical Division,
Group T-12, MS-J569, Los Alamos, NM 87545

1. INTRODUCTION

Development of the density functional theory (DFT) can be viewed as a sequence of three distinct steps. First, in their celebrated paper /1/ Hohenberg and Hohn have shown that the electronic wavefunction of the ground state is entirely determined by the corresponding density. This provided a *non-constructive proof* of the existence of the density functional for the

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

electronic energy, $E[\rho]$. It had taken 15 years before the next step was reached. The constrained search approach /2,3,4/ yielded an *implicit construction* for the functional $E[\rho]$. A rigorous prescription for $E[\rho]$ was given, but its algebraic form (either direct or indirect) remained unknown. Various approaches to this problem, often connected with the efforts to find a Schrödinger-like equation for the electron density, followed soon /5-10/. However, an *explicit construction* of $E[\rho]$ has become known only very recently /11-13/. In this approach, the M equations for the spin-orbitals, variational coefficients, etc. are replaced by M-1 equations for the auxiliary quantities and the equation for ρ itself. In principle, one can either solve or find approximate solutions to these auxiliary equations first and then solve the equation for ρ separately.

In this chapter we review this new approach to the construction of $E[\rho]$. The formalism is valid within any well-defined molecular orbital theory and provides means of reformulating it within the language of DFT.

2. MOLECULAR ORBITAL THEORIES

The molecular orbital theories, such as self-consistent field (SCF), configuration interaction (CI) and multi-configurational SCF (MCSCF), constitute minimization of the energy functional

$$E[\phi,\underline{c}] = \sum_{ij} A_{ij}(\underline{c}) \int \phi_i^*(\vec{r})\phi_j(\vec{r})\, v(\vec{r})\, d\vec{r}$$
$$+ (1/2) \sum_{ij} A_{ij}(\underline{c}) \int \nabla\phi_i^*(\vec{r})\cdot\nabla\phi_j(\vec{r})\, d\vec{r}$$
$$+ (1/2) \sum_{ijkl} A_{ij}(\underline{c})A_{kl}(\underline{c}) \iint \phi_i^*(\vec{r})\phi_j(\vec{r})G(\vec{r},\vec{r}')\phi_k^*(\vec{r}')\phi_l(\vec{r}')d\vec{r}d\vec{r}'$$

$$+ \sum_{ijkl} B_{ijkl}(\underline{c}) \iint \phi_i^*(\vec{r})\phi_j(\vec{r})G(\vec{r},\vec{r}')\phi_k^*(\vec{r}')\phi_l(\vec{r}')d\vec{r}d\vec{r}' \quad . \quad (1)$$

In the above, $G(\vec{r},\vec{r}') \equiv |\vec{r}-\vec{r}'|^{-1}$ for electrons, and $d\vec{r}$ takes care of both the Cartesian and spin spaces. The coupling tensors $\underline{A}$ and $\underline{B}$, specified in the particular theory, are functions of $N_c \geq 0$ variational parameters $\underline{c}$. The N_f one-particle functions (spin-orbitals) are subject to the orthonormality constraint

$$\int \phi_i^*(\vec{r})\phi_j(\vec{r})\, d\vec{r} = \delta_{ij} \quad , \quad 1 \leq i,j \leq N_f \quad . \quad (2)$$

For the system of N particles $N_f \geq N$, N_o ($N_f \geq N_o \geq 0$) of them are optimized. In particular, $N_o = N_f = N$, $N_c = 0$ for SCF; $N_o = 0$, $N_f > N$, $N_c > 0$ for CI; and $N_f \geq N_o > N$, $N_c > 0$ for MCSCF.

The particle density is given by

$$\rho(\vec{r}) = \sum_{ij} A_{ij}(\underline{c})\ \phi_i^*(\vec{r})\phi_j(\vec{r}) \quad . \quad (3)$$

Since ρ is real and non-negative everywhere, and integrates to N, the tensor $\underline{A}(\underline{c})$ is Hermitian and semi-positive definite, and its trace equals N. Using the definition (3) one can rewrite Eq.(1) as

$$\begin{aligned} E[\underline{\phi},\underline{c}] &= \int \rho(\vec{r})\ v(\vec{r})\ d\vec{r} + (1/2) \sum_{ij} A_{ij}(\underline{c}) \int \vec{\nabla}\phi_i^*(\vec{r})\cdot\vec{\nabla}\phi_j(\vec{r})\ d\vec{r} \\ &+ (1/2) \iint \rho(\vec{r})\ G(\vec{r},\vec{r}')\ \rho(\vec{r}')\ d\vec{r}\ d\vec{r}' \\ &+ \sum_{ijkl} B_{ijkl}(\underline{c}) \iint \phi_i^*(\vec{r})\phi_j(\vec{r})G(\vec{r},\vec{r}')\phi_k^*(\vec{r}')\phi_l(\vec{r}')\ d\vec{r}d\vec{r}' \quad . \quad (4) \end{aligned}$$

The energy E is a sum of four terms. The first one (the potential energy of particle-external potential interactions) and the third one (the direct part of the potential energy of particle-particle interactions) are sole functionals of ρ. The second one (the kinetic energy) and the fourth one (which describes both the exchange and correlation energies) depend also on vectors $\underline{\phi}$ and $\underline{c}$.

One can arrive at the sole density functional for E, valid within a particular molecular orbital theory, by minimizing (1) subject not only to the orthonormality conditions (2), but also to the constraint (3). If ρ denotes a fixed density, this can be written symbolically as

$$E[\rho] = \min_{(\underline{\phi}, \underline{c}) \to \rho} E[\underline{\phi}, \underline{c}] \quad . \tag{5}$$

Minimization of $E[\rho]$ yields the ground state density ρ_0 and the ground state energy $E[\rho_0]$.

The definition (5) constitutes the constrained search approach of Lieb /3/ and Levy /2/. Practical applications of (5) have been hindered for a long time because a construction of all pairs of vectors $(\underline{\phi}, \underline{c}) \to \rho$ have been unknown. This problem has been solved only very recently.

3. CONSTRUCTION OF ONE-PARTICLE WAVEFUNCTIONS

We start the construction of one-particle wavefunctions with the following

DEFINITION: A vector $\underline{\phi} \equiv \underline{\phi}(\vec{r})$ of N_f functions is called *proper* if:

1. The functions ϕ_i and ϕ_j are linearly independent for all $1 \leq i \neq j \leq N_f$.
2. $\sum_{i=1}^{N_f} \phi_i^*(\vec{r})\, \phi_i(\vec{r}) \neq 0$ for all $\vec{r} \in \mathbb{R}^3$.
3. There is a finite domain $\mathbb{D} \in \mathbb{R}^3$ such that $\prod_{i=1}^{N_f} \phi_i(\vec{r}) \neq 0$ for all $\vec{r} \in \mathbb{D}$.

For the quantum-mechanical applications it is also desirable that

each component of $\underline{\psi}$ is twice differentiable in $\mathbb{R}^3$.

The construction of all one-particle wavefunctions that conform to both (2) and (3) is provided by the following THEOREM 1 /11/: Let $\underline{\psi} \equiv \underline{\psi}(\vec{r})$ be a proper vector and $\rho \equiv \rho(\vec{r})$ be a given density which integrates to N. Let trace of $\underline{A}(\underline{c})$ is equal N. Let

$$\beta \equiv \beta(\vec{r}) = \sum_{ijkl} A_{ij}(\underline{c}) \; (S^{-1/2})_{ik} (S^{-1/2})^*_{jl} \; \psi^*_k(\vec{r}) \psi_l(\vec{r}) \quad , \tag{6}$$

where the generalized overlap matrix $\underline{S}$ is a Hermitian, positive-definite solution of the equation

$$S_{ij} = \int \psi^*_i \; \psi_j \; \beta^{-1} \; \rho \; d\vec{r} \quad . \tag{7}$$

1. The functions $\underline{\phi} \equiv \underline{\phi}(\vec{r})$ given by

$$\phi_i = \beta^{-1/2} \; \rho^{1/2} \sum_j (S^{-1/2})^*_{ij} \; \psi_j \quad , \tag{8}$$

conform to (2) and (3).

2. The above construction is invariant with respect to the transformation $\underline{\psi} \rightarrow \kappa \; \underline{\psi}$, where $\kappa \equiv \kappa(\vec{r})$ is any complex function with no zeros. One can therefore set $\psi_{N_f}(\vec{r}) \equiv 1$ without any loss of generality.

Inserting Eqs.(6) and (8) into (4) one obtains

$$E[\rho,\underline{\psi},\underline{c}] = V[\rho] + J[\rho] + T_w[\rho] + T_f[\rho,\underline{\psi},\underline{c}] + V_{xc}[\rho,\underline{\psi},\underline{c}] \tag{9}$$

where

$$V[\rho] = \int v(\vec{r}) \; \rho(\vec{r}) \; d\vec{r} \quad , \tag{10}$$

$$J[\rho] = (1/2) \iint \rho(\vec{r}) \; G(\vec{r},\vec{r}') \; \rho(\vec{r}') \; d\vec{r} \; d\vec{r}' \quad , \tag{11}$$

$$T_w[\rho] = (1/8) \int |\vec{\nabla}\rho(\vec{r})|^2 \rho^{-1}(\vec{r}) \, d\vec{r} \quad , \tag{12}$$

$$T_f[\rho,\Psi,\underline{c}] = (1/2) \sum_{ijkl} A_{ij}(\underline{c}) \, (S^{-1/2})_{ik} \, (S^{-1/2})^*_{jl}$$
$$\int \beta^{-1}(\vec{r}) \, \rho(\vec{r}) \, \vec{\nabla}\psi^*_k(\vec{r}) \cdot \vec{\nabla}\psi_l(\vec{r}) \, d\vec{r} - (1/8) \int \beta^{-2}(\vec{r}) \, \rho(\vec{r})$$
$$\left| \sum_{ijkl} A_{ij}(\underline{c}) \, (S^{-1/2})_{ik} \, (S^{-1/2})^*_{jl} \, \vec{\nabla}[\psi^*_k(\vec{r})\psi_l(\vec{r})] \right|^2 d\vec{r}, \tag{13}$$

$$V_{xc}[\rho,\Psi,\underline{c}] =$$
$$\sum_{ijkl} \sum_{pqrs} B_{ijkl}(\underline{c}) \, (S^{-1/2})_{ip} (S^{-1/2})^*_{jq} (S^{-1/2})_{kr} (S^{-1/2})^*_{ls}$$
$$\int\int \beta^{-1}(\vec{r})\beta^{-1}(\vec{r}')\rho(\vec{r})\rho(\vec{r}')\psi^*_p(\vec{r})\psi_q(\vec{r})G(\vec{r},\vec{r}')\psi^*_r(\vec{r}')\psi_s(\vec{r}')d\vec{r}d\vec{r}'. \tag{14}$$

Let us introduce the density amplitude

$$A \equiv A(\vec{r}) = \rho^{1/2}(\vec{r}) \quad . \tag{15}$$

Minimization of E yields the following pseudo-Schrödinger equation for A

$$[(-1/2) \, \vec{\nabla}^2 + V(\vec{r}) + V_{excl}(\vec{r})] \, A(\vec{r}) = \epsilon \, A(\vec{r}) \quad . \tag{16}$$

The potential V is a sum of the external potential v and the electrostatic potential due to ρ

$$V(\vec{r}) = v(\vec{r}) + \int \rho(\vec{r}') \, G(\vec{r},\vec{r}') \, d\vec{r}' \quad . \tag{17}$$

The *exclusion potential* V_{excl} /12/ arises from the Pauli exclusion principle. It is composed of the *Pauli potential* /8/

$$T(\vec{r}) = \frac{\delta \langle T_f \rangle}{\delta \, \rho(\vec{r})} \quad , \tag{18}$$

and the exchange-correlation potential

$$V_{xc}(\vec{r}) = \frac{\delta \langle V_{xc} \rangle}{\delta \rho(\vec{r})} \quad . \tag{19}$$

Both V and V_{excl} depend on the density ρ. Therefore Eq.(16) is non-linear and has to be solved iteratively. V_{excl} also depends on $N_f - 1$ auxiliary functions $\underline{\phi}$ and N_c coefficients $\underline{c}$. They can be obtained from the respective minimizations of E, or they can be assumed. This way one can obtain either exact or approximate V_{excl}.

The Lagrange multiplier ϵ allows for a proper normalization of A. It should not be confused with the chemical potential, since the latter one is obtained only by minimization with a variable number of particles /14/.

4. DENSITY DRIVEN SELF-CONSISTENT FIELD THEORY: FORMALISM

The above formalism can be illustrated by considering the density driven self-consistent field theory (DD-SCF) /12/. Obviously, it is an analog of the standard SCF approach. To obtain the working equations of DD-SCF one starts from Eq.(1) together with

$$A_{ij} = \delta_{ij} \quad , \quad 1 \leq i,j \leq N \quad , \tag{20}$$

$$B_{ijkl} = (-1/2)\ \delta_{il}\ \delta_{kl} \quad , \quad 1 \leq i,j,k,l \leq N \quad . \tag{21}$$

There are $N_f = N$ one-particle functions $\underline{\phi}$ and no ($N_c = 0$) parameters $\underline{c}$. Substantial simplifications occur: Eq.(6) becomes

$$\beta \equiv \beta(\vec{r}) = \sum_{ij} (S^{-1})^*_{ij}\ \phi^*_i(\vec{r})\phi_j(\vec{r}) \quad . \tag{22}$$

The fermionic kinetic energy reads

$$T_f[\rho,\underline{\psi}] = (1/2) \int \beta^{-1}(\vec{r})\ \rho(\vec{r}) \sum_{ij} (S^{-1})^*_{ij}\ \vec{\nabla}\psi^*_i(\vec{r})\cdot\vec{\nabla}\psi_j(\vec{r})\ d\vec{r}$$
$$- (1/8) \int \beta^{-2}(\vec{r})\rho(\vec{r})\ |\sum_{ij} (S^{-1})^*_{ij}\ \vec{\nabla}[\psi^*_i(\vec{r})\psi_j(\vec{r})]|^2\ d\vec{r} \quad . \quad (23)$$

V_{xc} consists only of the exchange term

$$V_{xc}[\rho,\underline{\psi}] = (-1/2) \times$$
$$\iint a(\vec{r},\vec{r}')a^*(\vec{r},\vec{r}')\beta^{-1}(\vec{r})\beta^{-1}(\vec{r}')\rho(\vec{r})\rho(\vec{r}')G(\vec{r},\vec{r}')\ d\vec{r}d\vec{r}' \ , (24)$$

where

$$a(\vec{r},\vec{r}') = \sum_{ij} (S^{-1})^*_{ij}\ \psi^*_i(\vec{r})\psi_j(\vec{r}') \quad . \quad (25)$$

There are two important theorems concerning the matrix $\underline{S}$ in the DD-SCF case /13/:

THEOREM 2: The matrix $\underline{S}$ exists for any N-dimensional proper vector $\underline{\psi}$ and any density ρ which integrates to N, and is defined up to a non-zero real multiplicative constant. The vector $\underline{\phi}$ is independent of this constant.

THEOREM 3: Let us define the following iteration

$$S^{(n+1)}_{ij} = \int \psi^*_i(\vec{r})\psi_j(\vec{r})\ [\sum_{kl} (S^{(n)-1})^*_{kl}\ \psi^*_k(\vec{r})\psi_l(\vec{r})]^{-1}\ \rho(\vec{r})\ d\vec{r} \ , \quad (26)$$

followed by normalization of $S^{(n+1)}$ (for example by forcing Tr $\underline{S}$ = 1). The iteration (26) converges linearly to $\underline{S}$ if the guess $\underline{S}^{(0)}$ is sufficiently close to $\underline{S}$.

In practical calculations we have found that, especially for large N, the convergence can be slow. However, application of the

matrix Aitken interpolation enhances greatly the rate of convergence /15/.

Let us define the tensor Ω as the solution of the following linear equation

$$\sum_{kl}\left\{ S_{km}S^{*}_{ln} + \kappa S_{kl}S^{*}_{mn} - \int \psi^{*}_{k}\psi_{l}\psi_{m}\psi^{*}_{n}\beta^{-2}\rho \, d\vec{r} \right\} \Omega_{kl,ij} = - \delta_{in}\delta_{jm} \tag{27}$$

where κ is an arbitrary non-zero real constant.

The explicit expressions for $T(\vec{r})$ and $V_{xc}(\vec{r})$ read /12/

$$\begin{aligned}
T(\vec{r}) &= [2\beta(\vec{r})]^{-1} \sum_{ij} (S^{-1})^{*}_{ij} \nabla\psi^{*}_{i}(\vec{r})\cdot\nabla\psi_{j}(\vec{r}) \\
&- [8\beta^{2}(\vec{r})]^{-1} \left|\sum_{ij} (S^{-1})^{*}_{ij} \nabla[\psi^{*}_{i}(\vec{r})\psi_{j}(\vec{r})]\right|^{2} \\
&+ [4\,\beta(\vec{r})]^{-1} \sum_{ij} \psi^{*}_{i}(\vec{r})\psi_{j}(\vec{r}) \sum_{kl} \Omega_{kl,ij} \times \\
&\Big[-2 \int \psi^{*}_{k}(\vec{r}')\psi_{l}(\vec{r}') \left[\sum_{mn} (S^{-1})^{*}_{mn} \nabla\psi^{*}_{m}(\vec{r}')\cdot\nabla\psi_{n}(\vec{r}')\right] \beta^{-2}(\vec{r}')\rho(\vec{r}')d\vec{r}' \\
&+ \int \psi^{*}_{k}(\vec{r}')\psi_{l}(\vec{r}') \left|\sum_{mn} (S^{-1})^{*}_{mn} \nabla[\psi^{*}_{m}(\vec{r}')\psi_{n}(\vec{r}')]\right|^{2} \beta^{-3}(\vec{r}')\rho(\vec{r}')d\vec{r}' \\
&+ 2 \int \left[\nabla\psi^{*}_{k}(\vec{r}')\cdot\nabla\psi_{l}(\vec{r}') \right] \beta^{-1}(\vec{r}') \, \rho(\vec{r}') \, d\vec{r}' \\
&- \int \nabla[\psi^{*}_{k}(\vec{r}')\psi_{l}(\vec{r}')]\cdot\Big\{ \sum_{mn} (S^{-1})^{*}_{mn} \nabla[\psi^{*}_{m}(\vec{r}')\psi_{n}(\vec{r}')] \Big\} \\
&\beta^{-2}(\vec{r}') \, \rho(\vec{r}') \, d\vec{r}'\Big]
\end{aligned} \tag{28}$$

and

$$\begin{aligned}
V_{xc}(\vec{r}) &= - \beta^{-1}(\vec{r}) \int a(\vec{r},\vec{r}')a^{*}(\vec{r},\vec{r}') \, \beta^{-1}(\vec{r}')\rho(\vec{r}') \, G(\vec{r},\vec{r}') \, d\vec{r}' \\
&+ \beta^{-1}(\vec{r}) \sum_{ij} \psi^{*}_{i}(\vec{r})\psi_{j}(\vec{r}) \sum_{kl} \Omega_{kl,ij} \times \\
&\int\int \left[\psi_{l}(\vec{r}') \, a(\vec{r}',\vec{r}'') \, \beta^{-1}(\vec{r}') - \psi_{l}(\vec{r}'') \right] \\
&\psi^{*}_{k}(\vec{r}')a^{*}(\vec{r}',\vec{r}'')\beta^{-1}(\vec{r}')\beta^{-1}(\vec{r}'')\rho(\vec{r}')\rho(\vec{r}'') \, G(\vec{r}',\vec{r}'') \, d\vec{r}'d\vec{r}'' \, .
\end{aligned} \tag{29}$$

The auxiliary functions ψ satisfy the equations

$$
\begin{aligned}
&- (1/2)\ \nabla^2\psi_\lambda(\vec{r}) - (1/2)\ \rho^{-1}(\vec{r})\ \nabla\psi_\lambda(\vec{r})\cdot\nabla\rho(\vec{r}) \\
&+ (1/2)\ \beta^{-1}(\vec{r})\ \nabla\psi_\lambda(\vec{r})\cdot \sum_{mn} (S^{-1})^*_{mn}\ \nabla[\psi^*_m(\vec{r})\psi_n(\vec{r})] \\
&- (1/4)\ \beta^{-2}(\vec{r})\ \psi_\lambda(\vec{r})\ \Big|\sum_{mn} (S^{-1})^*_{mn}\ \nabla[\psi^*_m(\vec{r})\psi_n(\vec{r})]\Big|^2 \\
&+ (1/4)\ \beta^{-1}(\vec{r})\psi_\lambda(\vec{r})\ \sum_{mn} (S^{-1})^*_{mn}\ [\psi_n(\vec{r})\ \nabla^2\psi^*_m(\vec{r}) + \psi^*_m(\vec{r})\ \nabla^2\psi_n(\vec{r})] \\
&+ (1/4)\ \beta^{-1}(\vec{r})\ \rho^{-1}(\vec{r})\ \psi_\lambda(\vec{r})\ \nabla\rho(\vec{r})\cdot \sum_{mn} (S^{-1})^*_{mn}\ \nabla[\psi^*_m(\vec{r})\psi_n(\vec{r})] \\
&- \int \psi_\lambda(\vec{r}')\ a^*(\vec{r},\vec{r}')\ \beta^{-1}(\vec{r}')\ \rho(\vec{r}')\ G(\vec{r},\vec{r}')\ d\vec{r}' \\
&+ \beta^{-1}(\vec{r})\ \psi_\lambda(\vec{r}) \int a(\vec{r},\vec{r}')\ a^*(\vec{r},\vec{r}')\ \beta^{-1}(\vec{r}')\ \rho(\vec{r}')\ G(\vec{r},\vec{r}')\ d\vec{r}' \\
&+ (1/4) \sum_{ij} \psi_j(\vec{r})\ [S^*_{\lambda i} - \beta^{-1}(\vec{r})\ \psi^*_i(\vec{r})\ \psi_\lambda(\vec{r})] \sum_{kl} \Omega_{kl,ij} \times \\
&\Big[- 2 \int \psi^*_k(\vec{r}')\psi_l(\vec{r}')[\sum_{mn} (S^{-1})^*_{mn}\nabla\psi^*_m(\vec{r}')\cdot\nabla\psi_n(\vec{r}')]\ \beta^{-2}(\vec{r}')\rho(\vec{r}')d\vec{r}' \\
&+ \int \psi^*_k(\vec{r}')\psi_l(\vec{r}')\ \Big|\sum_{mn} (S^{-1})^*_{mn}\ \nabla[\psi^*_m(\vec{r}')\psi_n(\vec{r}')]\Big|^2\ \beta^{-3}(\vec{r}')\rho(\vec{r}')d\vec{r}' \\
&+ 2 \int [\ \nabla\psi^*_k(\vec{r}')\cdot\nabla\psi_l(\vec{r}')\ ]\ \beta^{-1}(\vec{r}')\ \rho(\vec{r}')\ d\vec{r}' \\
&- \int \nabla[\psi^*_k(\vec{r}')\psi_l(\vec{r}')]\cdot\{\sum_{mn} (S^{-1})^*_{mn}\nabla[\psi^*_m(\vec{r}')\psi_n(\vec{r}')]\}\beta^{-2}(\vec{r}')\rho(\vec{r}')d\vec{r}' \\
&+ 4 \iint [\ \psi_l(\vec{r}')\ a(\vec{r}',\vec{r}'')\ \beta^{-1}(\vec{r}') - \psi_l(\vec{r}'')\ ] \\
&\psi^*_k(\vec{r}')\ a^*(\vec{r}',\vec{r}'')\ \beta^{-1}(\vec{r}')\beta^{-1}(\vec{r}'')\ \rho(\vec{r}')\rho(\vec{r}'')\ G(\vec{r}',\vec{r}'')\ d\vec{r}'d\vec{r}''\ \Big] \\
&= 0 \qquad\qquad \lambda = 1, \ldots, N-1\ , \\
&\psi_\lambda(\vec{r}) \equiv 1 \qquad\qquad \lambda = N\ . \qquad\qquad (30)
\end{aligned}
$$

5. DENSITY DRIVEN SELF-CONSISTENT FIELD METHOD: DISCUSSION

There are several available approaches to solving the DD-SCF problem:

A. SCF-like approach (not very elegant)

1. Guess Λ and ψ.

2. Calculate $\underline{S}$ and β (Eqs.7 and 22), α (Eq. 25), and Ω (Eq.27).
3. Calculate V (Eq.17), T (Eq.28) and V_{xc} (Eq.29).
4. Solve Eq.(16) for Λ and Eq.(30) for ψ.
5. Repeat steps 2-4 until self-consistence.

B. "True" DFT approach (rather futile)

1. For a given ρ, solve Eqs.(7, 22, 25, 27 and 30) analytically for ψ.
2. Substitute ψ into Eqs.(9, 23 and 24) to obtain $E[\rho]$ explicitly.
3. Obtain from $E[\rho]$ the variational equation for Λ and solve it self-consistently.

C. Fixed ψ approach (rather crude)

1. Assume ψ.
2. Guess Λ.
3. Calculate $\underline{S}$ and β (Eqs.7 and 22), α (Eq. 25), and Ω (Eq.27).
4. Calculate V (Eq.17), T (Eq.28) and V_{xc} (Eq.29).
5. Solve Eq.(16) for Λ.
6. Repeat steps 3-5 until self-consistence.

D. Hybrid approach

1. Assume a basis set for ψ.
2. Guess Λ.
3. Calculate $\underline{S}$ and β (Eqs.7 and 22), α (Eq. 25), and Ω (Eq.27).
4. Calculate V (Eq.17), T (Eq.28) and V_{xc} (Eq.29).
5. Solve Eq.(16) for Λ.
6. Minimize the energy (Eqs.9, 10, 11, 12, 23 and 24) with respect to ψ.
7. Repeat steps 3-6 until self-consistence.

The hybrid approach seems to be best suited for numerical computations. However, one should point out that even the fixed ψ approach is capable of providing reasonable results. For example, one can use the "exact" HF density for ρ and the functions R; R, x, y, z; and R, x, y, z, R^2, xR, yR, zR as the auxiliary functions for the Be, Ne and Ar atoms, respectively. The resultant values of the kinetic energy are 15.09 (Be), 133.8 (Ne) and 565.7 (Ar) compared to the exact values of 14.57, 128.6 and 526.8. This

example shows clearly that the results of DD-SCF calculations are expected to be quite insensitive to the choice of $\not{\phi}$.

The formalism of DD-SCF makes also possible to derive expressions for the Pauli and exchange potentials in terms of the Hartree-Fock spin-orbitals. The ith HF spin-orbital satisfies the equation

$$- (1/2)\ \nabla^2 \phi_i^{HF}(\vec{r}) + V(\vec{r})\ \phi_i^{HF}(\vec{r}) + (\hat{K}\phi_i^{HF})(\vec{r}) = e_i\ \phi_i^{HF}(\vec{r})\ , \quad (31)$$

where e_i is the ith spin-orbital energy and

$$(\hat{K}\phi_i^{HF})(\vec{r}) = - \sum_j \phi_j^{HF}(\vec{r}) \left[\int \phi_i^{HF}(\vec{r}')\phi_j^{HF*}(\vec{r}')\ G(\vec{r},\vec{r}')\ d\vec{r}' \right] , \quad (32)$$

defines the exchange operator. Using the HF spin-orbitals as the auxiliary functions in Eqs.(23) and (24) one obtains

$$\begin{aligned} T(\vec{r}) &= [2\ \rho^{HF}(\vec{r})]^{-1} \sum_i \left[\ \nabla\phi_i^{HF*}(\vec{r})\cdot\nabla\phi_i^{HF}(\vec{r})\ \right] \\ &- (1/8)\ [\rho^{HF}(\vec{r})]^{-2}\ |\nabla\rho^{HF}(\vec{r})|^2 - [\rho^{HF}(\vec{r})]^{-1} \sum_i e_i\ \phi_i^{HF*}(\vec{r})\phi_i^{HF}(\vec{r}) \\ &+ N^{-1} \sum_i e_i + [\rho^{HF}(\vec{r})]^{-1} \sum_{ij} \phi_i^{HF*}(\vec{r})\phi_j^{HF}(\vec{r}) \sum_{kl} \Omega_{kl,ij} \times \\ &\{ - \int (\hat{K}\phi_k^{HF})^*(\vec{r}')\phi_l^{HF}(\vec{r}')\ d\vec{r}' \\ &+ \int \phi_k^{HF*}(\vec{r}')\phi_l^{HF}(\vec{r}') \left[\sum_m (\hat{K}\phi_m^{HF})^*(\vec{r}')\phi_m^{HF}(\vec{r}')\right] [\rho^{HF}(\vec{r}')]^{-1} d\vec{r}'\ \} , \end{aligned} \quad (33)$$

and

$$\begin{aligned} V_{xc}(\vec{r}) &= [\rho^{HF}(\vec{r})]^{-1} \sum_i (\hat{K}\phi_i^{HF})^*(\vec{r})\phi_i^{HF}(\vec{r}) \\ &- [\rho^{HF}(\vec{r})]^{-1} \sum_{ij} \phi_i^{HF*}(\vec{r})\phi_j^{HF}(\vec{r}) \sum_{kl} \Omega_{kl,ij} \times \\ &\{ - \int (\hat{K}\phi_k^{HF})^*(\vec{r}')\phi_l(\vec{r}')\ d\vec{r}' \end{aligned}$$

$$+ \int \phi_k^{HF*}(\vec{r}')\phi_l^{HF}(\vec{r}') \left[\sum_m (\hat{K}\phi_m^{HF})^*(\vec{r}')\phi_m^{HF}(\vec{r}')\right] [\rho^{HF}(\vec{r}')]^{-1} d\vec{r}' \} . \tag{34}$$

The Lagrange multiplier turns out to be equal to the average spin-orbital energy

$$\epsilon = N^{-1} \sum_i e_i \quad . \tag{35}$$

Eqs. (33) and (34) can serve as a starting point for construction of approximate expressions for T and V_{xc}.

6. CONCLUSIONS

The formalism presented in the above paragraphs can be easily extended to other molecular orbital theories. One can obtain explicit expressions for the exclusion potential and the associated variational equations for the auxiliary functions $\underline{\phi}$ and the variational coefficients $\underline{c}$. This allows to approximate the exact pseudo-Schrodinger equation for the density amplitude up to any desired degree of accuracy. Since the density amplitude is a function of only three Cartesian coordinates, the equation can be solved numerically. The auxiliary functions can be approximated by finite basis sets. As demonstrated above, these bases can be of a significantly lower quality that the bases currently used in *ab initio* calculations.

Another possibility offered by the present formalism is the analysis of the results of molecular orbital calculations in terms of the Pauli and exchange-correlation potentials. Expressions analogous to Eqs.(33) and (34) can be set up for the natural orbitals. In principle, one can also study various correlation

problems such as the maximal fraction of the correlation energy that can be recovered within the particle density constrained to be equal to the HF one.

The present approach makes it possible to separate the computation of the particle density from the computation of spin-orbitals. Only one Lagrange multiplier is used. This allows for efficient numerical algorithms for solving the relevant variational equations. In this aspect, the present formalism differs significantly from the proposition of Nyden and Parr /10/, where N(N+1)/2 Lagrange multipliers have to be used for the N-particle problems.

ACKNOWLEDGMENT

This research was pursued under the auspices of U.S. D.O.E.

REFERENCES

1. P.Hohenberg and W.Kohn, Phys.Rev. B136, 864 (1964)
2. M.Levy, Proc.Natl.Acad.Sci. USA 76, 6062 (1979)
3. E.H.Lieb, Intern.J.Quant.Chem. 24, 243 (1983)
4. P.W.Payne, J.Chem.Phys. 71, 490 (1979)
5. G.Hunter, Intern.J.Quant.Chem. 9, 237 (1975); 19, 755 (1981); 29, 197 (1986)
6. M.Levy, J.P.Perdew and V.Sahni, Phys.Rev. A30, 2745 (1984)
7. E.N.Lesettre, J.Chem.Phys. 83, 1709 (1985)
8. N.H.March, Phys.Lett. 113A, 66 (1985) and the references cited therein
9. K.A.Dawson and N.H.March, J.Chem.Phys. 81, 5850 (1984)
10. M.R.Nyden and R.G.Parr, J.Chem.Phys. 78, 4044 (1983)
11. J.Cioslowski, Phys.Rev.Lett. 60, 2141 (1988)
12. J.Cioslowski, J.Chem.Phys. 89, 4871 (1988)
13. J.Cioslowski, Intern.J.Quant.Chem., S23, xxxx (1989)
14. R.G.Parr and L.J.Bartolotti, J.Chem.Phys. 78, 2810 (1983)
15. J.Cioslowski, J.Chem.Phys. 89, 2126 (1988)

THE GAUSSIAN-TYPE ORBITALS DENSITY-FUNCTIONAL APPROACH TO FINITE SYSTEMS

B. I. Dunlap

Theoretical Chemistry Section, Code 6119
Naval Research Laboratory
Washington, D.C. 20375-5000, U.S.A.

N. Rösch

Lehrstuhl für Theoretische Chemie
Technische Universität München
D-8046 Garching, Federal Republic of Germany

ABSTRACT

The linear combination of Gaussian-type orbitals (LCGTO) approach to Xα and density functional theory is reviewed, with particular emphasis on applications to large molecules and clusters. Fitting the potential is central to the LCGTO approach, and efficient and accurate ways to do so are described. Model cluster calculations apply these methods to the adsorption of alkali atoms and carbon monoxide on transition metal surfaces as well as the problem of CO vibrational shifts upon alkali coadsorption.

1. INTRODUCTION

In the twenty-five years since the Hohenberg-Kohn (HK) theorem /1/ suggested that something like Thomas-Fermi theory /2/, whose pivotal quantity is the electronic density and not the many-electron wavefunction of Schrödinger theory, allows an exact solution for the total energy of nondegenerate ground states of atoms, molecules, and solids, indisputable computational progress has been made in the application of density-based methods to these systems. Such methods are now collectively called density functional theory (DFT).

In the present work we would like to emphasize the computational perspective. From this point of view, theoretical progress in DFT is harder to define. This is partly due to the fact that one must often be ready to step outside the bounds of conventional DFT in order to achieve practical advances. It is possible that the empirical evidence gained from nearly forty years of experience with the Xα method /3,4/ will hold up, namely that it is best to compute a kinetic energy using quantum mechanical methods. If so, the Kohn-Sham (KS) extension /5/ of the HK theorem will remain important. The new concept added in the KS approach is one-electron orbitals, with which a kinetic energy

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

is evaluated quantum mechanically. In the KS formulation /5/ the orbitals are those of a noninteracting electron gas having the density of the real system. This may or may not be the optimal way to interpret the kinetic energy /6,7/ that is computed in what are now called local density functional (LDF) or KS methods.

The computational advantage of Slater's Xα approach /3,4/ over Hartree-Fock (HF) stems from from the fact that every orbital $\phi_i(\mathbf{r})$ is generated by the same (effective) real-space one-electron potential, V,

$$[-\tfrac{1}{2}\nabla^2 + V(\mathbf{r})]\,\phi_i(\mathbf{r}) = \epsilon_i\,\phi_i(\mathbf{r}). \qquad (1)$$

The potential of this homogeneous system of equations is called local. The HF equations are inhomogeneous and thus cannot be so written, because each orbital would then need a different $V(\mathbf{r})$. In this work we are interested in computationally efficient self-consistent methods based on this local potential approximation. For the most efficient SCF computations the local potential should be a spin density functional and not depend on any additional electronic degrees of freedom. We gain considerable computational advantage by avoiding methods that project out a component of the wavefunction, e.g. via angular momentum about some center as is common in pseudopotential methods /8/, or via each orbital as in HF /9/ or the self-interaction correction /10/. Due to the influence of KS, Slater's approach of separating the local potential V into parts originating from nuclear attraction and from interelectronic repulsion, and a remainder attributed to exchange (and correlation) effects, has became the conventional starting point in discussions of DFT.

Methods are called LDF if the exchange and correlation (XC) potential at each point in space is only a functional of the local density at that point. It is the local potential approximation, and not the LDF approximation, that assures computational efficiency. We note that the Coulomb potential, although not of the LDF form, does not present a major computational obstacle in DFT, due at least to the fact that it is a local potential. We would like to, but cannot define further the boundaries of computational efficiency within DFT. This review is about pushing back those boundaries in the Gaussian-type orbitals (GTO) approach to DFT.

Of local potential methods in current use, Xα is the simplest and oldest one. Such methods will continue to be used for some time despite the existence of more accurate quantum chemical methods, because calculations on larger systems are possible with simpler methods and there is a large body of computational experience using these methods. All other current LDF XC functionals are at least as complicated as that of Xα, which alone is quite a challenge in any analytic molecular orbital (MO) approach. When any LDF XC functional is treated numerically, however, there is essentially no difference in computational difficulty.

There is no guarantee that computationally efficient methods will remain accurate enough, but the advantages of local potential methods are substantial. As we shall see, local potential methods offer the hope of a first-principles N^3 quantum chemical method, N being the size of the orbital basis set. Challenges

to the KS formulation /6,7/ use symmetry-required degeneracy. Even if spin is neglected, symmetry properties of the electronic wavefunction,

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) ,$$

are lost in obtaining the density,

$$\rho(\mathbf{r}) = \gamma(\mathbf{r}, \mathbf{r}) \tag{2}$$

where

$$\gamma(\mathbf{r}, \mathbf{r}') = N \int \Psi^*(\mathbf{r}', \mathbf{r}_2, \ldots, \mathbf{r}_N)\, \Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)\, d\mathbf{r}_2 \ldots d\mathbf{r}_N .$$

Therefore no DFT can give the symmetry of the ground state, because it generates only the density. It is not clear whether or not local potential methods can correctly describe even the multiplet structure of atoms /11/, let alone the lowest state of each symmetry, as has been claimed /12/. In the approach of KS, what is needed is a connection between external fields, the local potential, acting on a noninteracting electron gas and symmetry or degeneracy of its corresponding interacting electronic wavefunction. Such a connection does not appear to exist /6,7/ if we require that the external fields of the interacting and noninteracting systems both have the same point-group symmetry.

Even if no connection at all can be made and we are forced to proceed via an alternative route in the case of degeneracy, local potential methods already require more information than straight DFT, namely the entire first order density matrix /13,14/, $\gamma(r, r')$, just as does HF. In contrast to HF, its off-diagonal part is used only to compute the kinetic energy. First-order density matrices can come from wavefunctions which in turn contain symmetry information. For example, it is common practice to use the square of the spin angular momentum operator to project out spin contamination in an unrestricted HF single-determinant solution /9/. While it is not clear whether or not an exact LDF theory exists, less accurate local potential methods can be made to have certain symmetries and thus have the corresponding degeneracies of the real systems /15,16/.

This work focuses on applications of local potential methods to intermediate-sized molecules and clusters (containing ten or more atoms), with an attempt to minimize the use of any further approximations. A decade ago it became possible to compute the total energy of diatomic molecules other than hydrogen within an LDF approximation accurately enough to demonstrate shortcomings of the models /17/. That method, which uses GTOs, has now been surpassed in accuracy by completely numerical methods for diatomic molecules /18-23/. These more accurate methods showed that the GTO binding energies for first row diatomics are accurate to within 0.1 eV /23/. Completely numerical methods for the near future appear limited to systems possessing axial symmetry, whereas the GTO approach has been applied to many large systems /24-31/.

The second section of this work describes how to variationally fit the charge density. This efficiently reduces the otherwise N^4 problem of computing the direct electron-electron repulsion to an N^3 problem, due to the special properties of the charge density. The third section discusses various ways to fit the complicated LDF expressions for XC on a three-dimensional molecular grid. The fourth section suggests a completely analytic treatment of the Xα functional, which allows exact analytic total energy gradients. The fifth section discusses SCF convergence and fractional occupation numbers (FON). The sixth section describes an application of these methods to a problem in surface science: the chemisorption of akali atoms and of carbon monoxide on nickel surfaces. The analysis of these systems allows us to understand effects observed when alkali metal atoms and carbon monoxide coadsorb on the same transition metal surface.

2. FITTING THE COULOMB POTENTIAL

Central quantities in local potential methods are the density of electrons of each spin direction and the total electronic density,

$$\rho(\mathbf{r}) = \rho_\uparrow(\mathbf{r}) + \rho_\downarrow(\mathbf{r}).$$

Here, up and down arrows refer to the two spin directions. For each spin orientation the density is expanded in terms of one-electron orbitals, e.g. for spin up

$$\rho_\uparrow(\mathbf{r}) = n_{i\uparrow}\, \phi^*_{i\uparrow}(\mathbf{r})\, \phi_{i\uparrow}(\mathbf{r}), \tag{3}$$

where $n_{i\uparrow}$ is the occupation number of the ith one-electron orbital. (We follow the convention that repeated indices in any product are to be summed over.) In nonrelativistic methods we seek stationary solutions to the energy functional,

$$\begin{aligned} E =& n_{i\uparrow} \int \phi_{i\uparrow}(\mathbf{r})\, h_{op}\, \phi_{i\uparrow}(\mathbf{r})\, d\mathbf{r} + n_{i\downarrow} \int \phi_{i\downarrow}(\mathbf{r})\, h_{op}\, \phi_{i\downarrow}(\mathbf{r})\, d\mathbf{r} \\ &+ E_c(\rho) + E_{xc}(\rho_\uparrow, \rho_\downarrow) + E_{nuc}, \end{aligned} \tag{4}$$

where h_{op} is the sum of the usual kinetic energy and the nuclear attraction operator, E_{nuc} is the nuclear-nuclear replusion term, and E_c and E_{xc} are density functional expressions for the self-Coulomb energy of a charge density ρ and for the XC energy of charge densities $\rho_\uparrow$ and $\rho_\downarrow$. The latter two terms are computationally the most challenging terms in the local potential approach. The first is discussed in this section, and the second in the next two sections.

The DFT expression for the electron-electron Coulomb repulsion is that of a classical charge density ρ,

$$E_c = \tfrac{1}{2}[\rho\|\rho], \tag{5}$$

where

$$[f\|g] = \int d\mathbf{r}_1 d\mathbf{r}_2\, f(\mathbf{r}_1)\, g(\mathbf{r}_2)\, /\, r_{12}.$$

Substituting the local potential expression for the density, Eq. 3, into Eq. 5, we see that the electron-electron Coulomb repulsion problem contains terms involving four atomic centers in the linear combination of atomic orbitals (LCAO) approach. The direct LCGTO approach to any local potential method will therefore result in an N^4 computation, as does the LCGTO approach to HF. (For special cases, such as when the size of the molecule becomes much larger than the spatial extent of the most diffuse GTO basis function, HF and direct local potential calculations grow less rapidly than N^4. Such considerations simplify all GTO calculations to more or less the same degree, and thus do not affect this discussion.) If we do not use some kind of approximation here, we lose all the practical advantages of the local potential approach when we use Gaussian basis sets. Therefore, fitting E_c seems unavoidable /32/. Furthermore, if we want to use incomplete basis sets the problem must afford a variational treatment.

An important aspect of most GTO *ab initio* quantum chemical methods in use today is that they are variational in the total energy. Therefore relatively inaccurate (in a simple least-squares sense /33/), but systematically related wavefunctions can yield rather accurate total energy differences and thus reliable molecular potential energy surfaces. Of course, in chemistry and surface science we are interested in many properties other than the total energy, such as electronic excitations. The point is that if practicality dictates that we must use incomplete basis sets, then it is extremely valuable to use some sort of variational principle. At this point the leading all-purpose candidate involves the total energy, simply because of its central role both in the *ab initio* and the DFT formalisms. Even for cases such as computing hyperfine coupling constants where a variational total energy seems least appropriate, the energy variational principle can be made to suffice /34/, which is good because such calculations are usually only meaningful around the relevant minimum of the molecular potential energy surface.

No variational local potential molecular method now exists due to the complicated functional dependences on the density of current XC functionals; for example, when a grid of points of is used to treat these XC functionals numerically, the location of each point is never altered during the SCF process. Variational methods are possible for at least some LDF E_{xc} expressions /35/, and it is quite possible that problematical LDF XC functionals can be reparametrized using simpler functional forms.

These two necessary conditions, fitting and doing so variationally, are met in the GTO approach using the method of Dunlap, Connolly, and Sabin /17/. To review this approach, we write the total energy as,

$$E = E_{kinetic} + E_{potential} = T(\rho) + \int \rho(\mathbf{r})\, U(\rho(\mathbf{r}))\, d\mathbf{r}.$$

Here $T(\rho)$ is the local potential expression for the kinetic energy and varying this expression, written symbolically,

$$V(\rho) = \frac{\delta\,[\rho U(\rho)]}{\delta\rho} \tag{6}$$

gives the effective one-electron local potential for Eq. 1. To consider the effect of fitting on the total energy, we denote fitted quantities by drawing a line over them. If the potential is fit, then the total energy should be evaluated,

$$E - T(\rho) \approx \int \rho(\mathbf{r})\,\overline{V}(\rho(\mathbf{r}))\,d\mathbf{r} + \int \overline{\rho(\mathbf{r})\,[U(\rho(\mathbf{r})) - V(\rho(\mathbf{r}))]}\,d\mathbf{r}\,, \qquad (7)$$

because then the nonfitted density only occurs multiplied by the local potential. Variation with respect to the orbitals (occurring only in the nonfitted density) gives the local potential term of the one-electron equation (1), therefore the energy of Eq. 7 is insensitive in first order to changes of the one-electron orbitals. If, furthermore, variation of this expression with respect to the fitted quantities is stable, then we have a viable method that is insensitive to all first order errors. This important result can also be applied to any part of the potential, as we shall see in the following and the next two sections.

Applied to fitting the direct electronic Coulomb potential, Eq. 7 gives the expression /17/,

$$E_c \approx [\rho \,\|\, \overline{\rho}] - \tfrac{1}{2}[\overline{\rho} \,\|\, \overline{\rho}]\,. \qquad (8)$$

Fortunately varying this expression with respect to ρ is equivalent to varying the functional,

$$\Delta E_c = \tfrac{1}{2}[\rho - \overline{\rho} \,\|\, \rho - \overline{\rho}] \geq 0,$$

which, through Poisson's equation, is equivalent to least-squares fitting the electric field generated by the electrons /17/,

$$\Delta E_c \propto \int |\,\nabla U_c - \nabla \overline{U}_c\,|^2\,d\mathbf{r}.$$

Having discovered the only stationary way to fit the charge density that is also insensitive to first order changes in the one-electron orbitals, we must address the problem of practicality, i.e. we must improve upon the N^4 direct approach. Simple physical arguments ensure that this is possible. Because the orbital basis set at hand is reasonably complete for treating a large number of functions, namely the collection of one-electron orbitals, one might expect it to do pretty well in fitting one more similar, but particularly nice function, the LCGTO density generated by the LCGTO orbitals. This quantity has three features that make it easier to deal with than the exact orbitals. First, because the product of two Gaussians is a third Gaussian, the LCGTO density is exactly expressible as a finite sum of Gaussians. Second, the LCGTO density is cuspless. Third, the LCGTO density is nodeless.

A potential problem arises in fitting the LCGTO density due to the fact that multiplying two atomic orbitals of high angular momentum together gives terms having the combined, and thus quite high, angular momentum. This does not seem to be a problem in practice, however, as is suggested by a consideration of the highly successful LCAO approach itself. Good basis functions for the MO basis are the solutions of the atomic problem, augmented usually only in the valence region. These atomic functions are generated using the

'central-field approximation'. In local potential methods this approximation is to neglect all but zero angular momentum contributions to the potential. A direct extension of this LCAO approach to fitting the potential suggests including (perhaps in segments) the spherically symmetric central-field potentials of each atom as the zeroth-order basis for fitting the potential. As implemented in either the muffin-tin /4/ or self-consistent-charge /36/ approximations, such an approach yields a simple and rather reliable description of molecules. Of course, to get the most accurate results, one must augment this zeroth-order basis with either off-center or nonzero angular momentum functions.

For large systems we have found that it almost always suffices to include three subbases for each atom: s-type, r^2-type, and one with the lowest symmetry-allowed angular momentum for that atom in its environment. The s-type subbasis is generated from the s-type orbital basis with each exponent doubled. The r^2-type subbasis arises from the spherically symmetric part of the product of two p-type orbital functions. If a few of them are used to reproduce the shoulders of the atomic density due to core shell structure, then the core s-type fitting functions turn out to have positive cofficients in every molecular environment. That enhances numerical stability. We choose three to five even-tempered exponents with the central one being close to the inverse square of the atomic radius for our nonzero angular subbasis /31,37/. The next highest symmetry-allowed angular momentum and off-center s-type basis functions are often used to test effective completeness /31,37/. So described, the charge density basis sets that we employ are clearly smaller than the orbital basis set and we have an N^3 method for treating E_c unless the orbital basis set is heavily contracted. Contraction of the core fitting functions is useful, but for some reason cannot be used in the valence region /38/.

Is fitting the charge density or something similar useful to other quantum chemical methods? The first order density matrix in the GTO approach is more complicated than the density because it is a function of two positional variables, not one like the density. Therefore direct extension of this approach to fitting the HF exchange energy, through fitting the first order density matrix, is not as effective as is fitting E_c through fitting the density /39/. All direct and exchange two-electron integrals are four-center Coulomb integrals, however. The set of four-center integrals can, in turn, be thought of as direct integrals, i.e. a product of two orbitals at one position interacting with a product at another position. Now we can turn the above logic around. If the density can be fit within a given level of accuracy using a small subset of these pairs of functions, then a large subset of all pairs of orbital functions must be redundant in describing the Coulombic interaction – within the very same level of accuracy. This effective linear dependence shows up as small eigenvalues in the diagonalization of the set of all four-center Coulomb integrals. By eliminating those components of this product basis that give rise to eigenvalues below a realistic tolerance, significant savings, albeit using a somewhat complicated algorithm, are possible in the *ab initio* approach to quantum chemistry /40,41/. Fitting the charge density variationally can be done for infinite systems, provided that the charge density of the electrons and nuclei are both fit to charge-neutral functions and care is taken to remove the self-Coulomb interaction of each nucleus /42,43/.

3. THE GRID PROBLEM

All current LDF implementations for finite systems employ a grid of points somehow spread throughout the molecule. On the other hand, until recently /33/ no *ab initio* GTO method used a three dimensional grid of points. The reason for this difference is that LDF XC expressions,

$$E_{xc} = E_{xc}(\rho_\uparrow(\mathbf{r}), \rho_\downarrow(\mathbf{r})),$$

are quite complicated functions of the two spin densities, $\rho_\uparrow(\mathbf{r})$ and $\rho_\downarrow(\mathbf{r})$, to date requiring yet another basis to fit the related XC densities over this grid. For Xα /4/, the XC energy is the integral of the 4/3 power of these spin densities,

$$E_{xc} = C_\alpha [\rho_\uparrow^{4/3}] + C_\alpha [\rho_\downarrow^{4/3}] \tag{9}$$

where

$$[f] = \int f(\mathbf{r})\, d\mathbf{r},$$

and

$$C_\alpha = -\alpha \tfrac{9}{8} \left(\tfrac{3}{\pi}\right)^{1/3}.$$

The Xα XC energy and potential densities are proportional,

$$\begin{aligned} {}^\alpha U_{xc}^\uparrow &= C_\alpha \rho_\uparrow^{1/3} \\ {}^\alpha V_{xc}^\uparrow &= \tfrac{4}{3} C_\alpha \rho_\uparrow^{1/3}, \end{aligned} \tag{10}$$

affording simultaneous treatment of these two quantities which, in general, require separate numerical fits and could potentially require separate numerical procedures.

Important alternatives to Xα are the LDF XC parametrizations /44,45/ that are based on the essentially exact Monte-Carlo results of Ceperley and Alder /46/ for the interacting electron gas. These calculations can be performed in the limits of no or full spin polarization /46/. In order to represent concretely these more complicated LDF XC functionals, we give the parametrization of Perdew and Zunger /44/ (PZ), which we have used in our LCGTO work. Apart from its dependence on spin polarization, it is practically exact within the LDF XC approximation due to the quality of the fit. The PZ energy densities are defined in terms of r_s, which is the radius of a sphere containing the unit charge and thus is implicitly defined by $\rho^{-1} = 4\pi r_s^3/3$. For r_s greater than unity, the PZ parameterization of the XC energy density is

$${}^{PZ}U_{xc} = {}^{2/3}U_{xc} + \frac{\gamma}{1 + \beta_1 \sqrt{r_s} + \beta_2 r_s} \tag{11}$$

with $\gamma = -0.1423$ (-0.0843), $\beta_1 = 1.0529$ (1.3981), and $\beta_2 = 0.3334$ (0.2611) in the unpolarized (completely spin-polarized) limit. For r_s less than than unity, the PZ parameterization of the XC energy density is

$${}^{PZ}U_{xc} = {}^{2/3}U_{xc} + A \ln r_s + B + C r_s \ln r_s + D r_s \tag{12}$$

with $A = 0.033$ (-0.01555), $B = -0.048$ (-0.0269) $C = 0.0020$ (0.0007), and $D = -0.0116$ (-0.0048) in the unpolarized (completely spin-polarized) limit. In these equations, exchange is treated via Xα using the variational choice /5,47/ of $\alpha = 2/3$. The PZ equations for V_{xc}, from Eq. 6, are even more complicated,

$$^{PZ}V_{xc} = {}^{PZ}U_{xc} - \frac{r_s}{3}\frac{d\,{}^{PZ}U_{xc}}{dr_s}.$$

For intermediate amounts of spin polarization, the PZ XC energy and potential densities are interpolated by forcing the ratio of corresponding quantities for the two spin directions to be the same as in Xα. There is little difference between these two LDF XC functionals, either for $\alpha = 2/3$ or for $\alpha = 0.7$ that we, as well as others /48/, use. The spin-dependence of Xα is too strong /49/, much like that of HF (the approximation of which being Slater's original goal /4/ when deriving the Xα potential). There is a greater tendency to overbind nonradical (nonspin-polarized) molecules in PZ, because a monotonically increasing function of the density is added to the Xα form of the XC energy.

Ideally, the equations to be solved in fitting these expressions should follow variationally from Eq. 7. Due to the appearance of the 1/3 and 1/6 power of the density as well as its logarithm, however, these LDF XC expressions are difficult to treat analytically in the LCGTO approach. If they cannot be treated analytically, we must numerically integrate at least the XC component of the completely approximated third term on the right-hand side of Eq. 7.

At this point we reach a quandary due to the fact that in general no chemically accurate three-dimensional numerical integration scheme exists (although direct integration schemes are useful /50-52/). If the one-electron XC potential, V_{xc}, is fit to a combination of analytic functions then the second term,

$$\int \rho(\mathbf{r})\,\overline{V}_{xc}(\rho(\mathbf{r}))\,d\mathbf{r},$$

can be more accurately integrated analytically than numerically. Thus one might argue that it should be done analytically /32,48/, a choice that we also make. Nevertheless, Eq. 7 suggests the important result that the weight function, $w(\mathbf{r})$, in the numerical approach to the least-squares fitting problem,

$$\delta \int |V_{xc}(\mathbf{r}) - \overline{V}_{xc}(\mathbf{r})|^2\, w(\mathbf{r})\,d\mathbf{r} = 0, \tag{13}$$

should be the quantity /53/,

$$w(\mathbf{r}) = \rho(\mathbf{r}) / \overline{V}_{xc}(\mathbf{r}),$$

provided $\overline{V}$ (dropping the subscript) is nodeless, like V, otherwise it is adviseable to use V in this weight. This choice ensures, numerically at least, that the two integrals involving $\overline{V}$ and V respectively cancel in Eq. 7. This

leaves only a term involving U as should be the case, because it alone is the energy density. At this point we again face the analytic vs. numerical predicament. Perhaps the expressions are more accurate if U is fit to a sum of analytic functions rather than straight numerically integrated. We choose to fit.

Having determined what to fit, we must determine how to optimally choose the smallest number of points, because with efficient analytic integral evaluation this is the most time consuming part of the calculation. From research on the atomic problem, we know how to optimally choose radial points about each atom, at least for treating the full potential (if pseudopotentials are used, more uniform and hence less problematical distributions can be used /54/). The radial mesh should be logarithmic /55/, and start with the first Herman-Skillman radial step size /56/,

$$\Delta r_n = \Delta r_0 \, 2^{n/N} \qquad \text{where} \qquad \Delta r_0 \equiv 0.0025 \left(\frac{9\pi^2}{128Z}\right)^{1/3} \text{Bohr},$$

and N is the number of radial step it takes for the step size to double. We normally choose N = 40. For angular points optimal choices can be found in tables /57/, but they must be randomly rotated from one radial shell to next one to avoid biasing certain directions /58/ (in the molecular environment only). Increased efficiency results from reducing the number of angular points in core regions /58/. Decreasing the number of radial points in the valence region is not as advantageous due to the logarithmically increasing radial mesh spacing.

When atomic point distributions are combined in the molecular environment, a final problem with numerical treatments of the XC potential arises, namely that the atomic grids overlap, and thus the weights or volume elements for each point must change. This is probably the most serious numerical problem, at least to consistently treating different molecules, conformational isomers or energy hypersurfaces in general. The safest way proceeds by using analytic formulae to reduce the weight of a point asymptotically to zero as it approaches a foreign nucleus. This is the strategy followed in the DVM approach /48,59/. However that method leads to the largest number of points. Alternatively, one can exclude the region of space nearest a foreign atom, defining Wigner-Seitz cells or Voronoi polyhedra. Direct application of this method leads to discontinuous potential surfaces as a point crosses a cell boundary /60/. We linearly interpolate from one to zero over the current radial step size, higher-order interpolating polynomials have no noticeable effect /61/. Alternatively, the point problem may be minimized by adding and substracting quantities of similar magnitude to compute only the much smaller differences. This approach /62/ has recently culminated in what appears to be a practical first derivative method for the DVM method /63,64/. Numerical second order gradients are possible for small systems /33/.

4. A STATIONARY METHOD WITH ANALYTIC GRADIENTS

In the second section we used Eq. 7 to variationally fit the charge density. In the third section we used it to determine weights for numerically fitting V_{xc}. Now we use it to analytically fit ${}^{\alpha}V_{xc}$, yielding an analytic and variational

LCGTO Xα method /17,65/ and its energy gradients /35/. In the interest of saving space, we deal with the total energy, rather than paralleling the previous two sections. Like the basic LCGTO Xα approach of Sambe and Felton /32/, the method requires two auxiliary basis sets. With the first Gaussian basis set we fit the charge density,

$$\rho(\mathbf{r}) \approx \overline{\rho}(\mathbf{r}) = \sum \rho_i f_i(\mathbf{r}),$$

and with the second auxiliary basis we fit the XC energy density for each spin,

$$U_{xc}(\mathbf{r}) \approx \overline{U}_{xc}(\mathbf{r}) = \sum x_i g_i(\mathbf{r}),$$

(For Xα the fit to the potential density is proportional to this fit.)

Directly applying Eq. 7 for each of these two approximations gives the analytic LCGTO Xα total energy /17/:

$$\begin{aligned} E = {} & n_{m\uparrow}\phi^{m*}_{i\uparrow}\phi^{m}_{j\uparrow}\left\{h_{ij} + \rho_k\,[ij\|k] + \tfrac{4}{3}C_\alpha\, x^{\uparrow}_k\,[ijk]\right\} \\ & + n_{m\downarrow}\phi^{m*}_{i\downarrow}\phi^{m}_{j\downarrow}\left\{h_{ij} + \rho_k\,[ij\|k] + \tfrac{4}{3}C_\alpha\, x^{\downarrow}_k\,[ijk]\right\} \\ & - \tfrac{1}{2}\,\rho_i\rho_j[i\|j] - \tfrac{1}{3}\,C_\alpha\,(x^{\uparrow}_i x^{\uparrow}_j x^{\uparrow}_k x^{\uparrow}_l + x^{\downarrow}_i x^{\downarrow}_j x^{\downarrow}_k x^{\downarrow}_l)\,[ijkl] + E_{nuc}, \end{aligned} \tag{14}$$

where m refers to the mth one-electron orbital. In each primitive matrix element, the index i, j, k, or l points to the correspondingly numbered function from the basis set appropriate to the linear coefficient in question. In this equation the linear coefficients $\{\phi_i\}$ expand the set of orbitals $\{\phi_m(\mathbf{r})\}$ within the appropriate primitive basis set. The coefficients $\{\rho_i\}$ expand the charge density. The coefficients $\{x_i\}$, in turn, expand its cube root, $x(\mathbf{r}) = \rho(\mathbf{r})^{1/3}$. In our fit to the XC potential, all nonzero-angular-momentum Gaussian exponents are the same as those used in fitting the density ρ, but the zero-angular-momentum exponents are reduced by a factor of three, for obvious reasons. Because the cube root of the charge density is a much slower varying function of position than the density itself, fewer primitive functions are necessary. In fact, transition metal atoms have positive core x_i coefficients without including any r^2-type basis functions.

All of the linear coefficients in Eq. 14 are determined by making this expression stationary, subject to the appropriate constraints, obligatory orthonormality constraints on the KS orbitals and any precautionary constraints that one might want to impose on the charge density and XC fits. Including these constraints is equivalent to an unrestricted variation of a Lagrangian, for example /35/,

$$\begin{aligned} L = {} & E - n^{\uparrow}_m \epsilon^{\uparrow}_m \{\phi_{i\uparrow}\,\phi_{j\uparrow}\,[ij] - \delta_{ij}\} - n^{\downarrow}_m \epsilon^{\downarrow}_m \{\phi_{i\downarrow}\,\phi_{j\downarrow}\,[ij] - \delta_{ij}\} \\ & - \epsilon_0\,\{\rho_i\,[i] - N\}, \end{aligned} \tag{15}$$

where the various quantities ϵ allow satisfaction of the constraints they multiply. As a precautionary measure, we require the integral of the charge density

to equal the number of electrons. This constraint, represented by ϵ_0, is not essential, though /66,67/. Even for large systems its effect on the total energy is negligible, but it is useful because the size of its corresponding Lagrange multiplier indicates the quality of the fit /60/.

The problem with Eq. 14 is that it requires N^4 integrals, which, fortunately, can be avoided /35,65/. Suppose that we somehow have fit $x(\mathbf{r})$, the cube root of the charge density to a sum of Gaussians. Consider the square of that fit, $y(\mathbf{r}) = \overline{x}^2(\mathbf{r})$, and a fit to it $\overline{y}(\mathbf{r})$. We now have exactly the same situation that we had in fitting the charge density. Thus y is nodeless, cuspless, and has a finite expansion in terms of Gaussians, and again we have reduced the order of the calculation. Specifically,

$$x_i x_j x_k x_l [ijkl] = x_i x_j M_{mij}\, x_k x_l M_{mkl}, \tag{16}$$

where

$$M_{mij} = S^{-1/2}_{ml}\,[lij], \tag{17}$$

and $S^{-1/2}$ is the appropriate matrix to Löwdin orthogonalize the positive-definite overlap matrix in the basis set appropriate to y. This amounts to formally introducing the resolution of the identity into the matrix $[ijkl]$ which is correct under the assumption that the y basis is complete enough to span all appropriate products in the x basis.

The analytic expression for first-order gradients follow directly upon noticing that whenever we have a solution of Eq. 15, L equals E, because the constraints are equality constraints. Therefore the gradients of L and E are identical for all stationary solutions. The first partial derivative of L with respect to any linear basis set expansion coefficient ($\{\phi_i\}$, $\{\rho_i\}$, or $\{x_i\}$) is zero. Thus the gradients of the LCGTO Xα total energy with respect to the position of nucleus A includes only the gradients of the primitive integrals in Eq. 15:

$$\begin{aligned}
\nabla_A E = {} & P^{\uparrow}_{ji}\left\{\nabla_A h_{ij} + \rho_k \nabla_A [ij\|k] + \tfrac{4}{3} C_\alpha\, x^{\uparrow}_k \nabla_A [ijk]\right\} \\
& + P^{\downarrow}_{ji}\left\{\nabla_A h_{ij} + \rho_k \nabla_A [ij\|k] + \tfrac{4}{3} C_\alpha\, x^{\downarrow}_k \nabla_A [ijk]\right\} \\
& - \{\epsilon P^{\uparrow}_{ji} + \epsilon P^{\downarrow}_{ji}\}\nabla_A [ij] - \tfrac{1}{2}\rho_i \rho_j \nabla_A [i\|j] \\
& - \tfrac{1}{3} C_\alpha (x^{\uparrow}_i x^{\uparrow}_j x^{\uparrow}_k x^{\uparrow}_l + x^{\downarrow}_i x^{\downarrow}_j x^{\downarrow}_k x^{\downarrow}_l)\nabla_A [ijkl] + \nabla_A E_{nuc},
\end{aligned} \tag{18}$$

where $\epsilon P^{\uparrow}_{ji} = \epsilon^{\uparrow}_m n^{\uparrow}_m \phi^{m*}_{i\uparrow} \phi^{m}_{j\uparrow}$ is the one-electron eigenvalue-weighted density matrix for spin up.

An analytic approach is only presently available for Xα. Most local potential methods simply add a density functional expression for "correlation" to Xα, $\alpha = 2/3$, however. If these relatively small additions have significantly slower spatial variation than Xα, then the "exchange only" part should be treated analytically. In this case, both energies and gradients would be less susceptible to errors due to using a finite number of points.

5. SCF CONVERGENCE AND FON

Local potential DFT calculations on large systems are quite demanding because an orbital for each electron must be explicitly computed. Furthermore, SCF convergence is slowed by a small gap at the Fermi energy which is the average energy between the highest occupied MO (HOMO) and the lowest unoccupied MO (LUMO) for systems with integral occupation numbers. Such a small gap is often accompanied by another problematical feature, a high density of one-electron energies, density of states (DOS), near the Fermi energy. As the system grows in size both problems may worsen, particularly if units having levels in this region are added to the system, such as increasing the number of transition metal atoms modelling a surface in chemisorption problems. The gap at the Fermi energy is, as a rule, smaller (and concomitantly the DOS in its vicinity higher) for methods based on Eq. 1 than for HF. In HF each occupied orbital interacts only with the remaining occupied orbitals, whereas the unoccupied orbitals are repelled by *all* the occupied orbitals widening the gap, in contrast to local potential methods.

To see how this reduced gap affects SCF convergence, we must compute a correction to Eq. 7. That equation is appropriate for a fixed density, but in the SCF process if the fit to the potential changes, by $\Delta\overline{V}$, then the electronic density changes, too. The first order change in the energy is included in Eq. 7, but the second order change,

$$E_2 = \frac{n_i - n_j}{\epsilon_i - \epsilon_j} |\langle i|\Delta\overline{V}|j\rangle|^2, \tag{19}$$

an expression familiar from second order perturbation theory, is not. Here, $\langle i|\Delta\overline{V}|j\rangle$ is the one-electron matrix element of the change in the potential. In this and the following equation the symbol Δ is used to signify the difference between the potential, or density, at the beginning of an iteration in which the one-eigenvalue problem is solved (requiring some potential), and the potential that is necessary to move toward self-consistency. Each term in this expression lowers the energy (for ground states) unless the i and j occupation numbers are the same or if the change in the potential does not mix the two states, in which case it vanishes. For finite systems, $\langle i|\Delta\overline{V}|j\rangle$ will not vanish unless required by symmetry, and levels cannot become degenerate unless they are of different symmetry. Therefore, this expression can be used to obtain quadratic SCF convergence /68/. If the XC functional is handled numerically, however, the problems with equating numeric and analytic integrals, discussed in section 3, arise and Eq. 19 cannot be used effectively. Instead one can use it to treat the largest and most rapidly varying part of the potential that can vary in the SCF process, the direct Coulomb repulsion. In this case Eq. 8 must be modified according to

$$[\overline{\rho}\,\|\overline{\rho}] \longrightarrow [\overline{\rho} + \Delta\overline{\rho}\,\|\overline{\rho} + \Delta\overline{\rho}] + 2\frac{n_i - n_j}{\epsilon_i - \epsilon_j}\left|[\Delta\overline{\rho}\|ij]\right|^2. \tag{20}$$

For small systems the summation over i and j can be limited to a few levels around HOMO and LUMO where the energy denominators are smallest /69/.

For large systems, unfortunately, treating only a part of the potential, limiting this summation, or monitoring the size and sign of $\Delta\overline{V}$ from iteration to iteration /70-72/ is problematical.

If symmetry restrictions are imposed on $\overline{V}$ for finite systems, then one-electron orbitals can be degenerate. In fact, for symmetric molecules, there is often some geometry at which two levels of different symmetry will become degenerate at the Fermi energy. If there is degeneracy at the Fermi energy, one can choose to occupy the orbitals equally /73/ (resulting in FON) or not /74/. In such a case, yet another, lower energy solution of Eq. 4 can be found by breaking symmetry /15/. The energy can be lowered still further by symmetrizing the single-determinant wavefunction used to compute the kinetic energy /75/. These broken-symmetry solutions to the degeneracy problem other than FON are computationally intensive, however. A serious drawback of FON is the lack of a unique wavefunction to associate with the calculation. This shows up, for example, in our inability to consistently compute energies and matrix elements for optical transitions involving orbitals that are partially occupied in the FON 'ground state'.

Primarily for practical reasons we choose to invoke symmetry on the one-electron orbitals for our larger clusters. This saves a considerable amount of computer time. A practical way to ensure that we stay on the 'ground state' manifold of our potential surfaces is to use FON that are a smooth function of nuclear position. For gas phase clusters we prefer to use FON that are given by the Fermi function,

$$n_i(\epsilon_f) = m_i \left[1 + \exp\left(\frac{\epsilon_i - \epsilon_f}{\sigma}\right)\right]^{-1}, \tag{21}$$

where m_i designates the degeneracy of level i. The Fermi energy ϵ_f must be found self-consistently given the total number of electrons, and the width, σ, which is perhaps best chosen to be room temperature, 0.025 eV. For surface calculations we treat the embedding problem by formally broadening the cluster levels and filling the resulting formal density of one-electron levels from below /28,31/,

$$n_i(\epsilon_f) = m_i \int_{-\infty}^{\epsilon_f} \exp\left[-\frac{(\epsilon - \epsilon_i)^2}{2\sigma^2}\right] \frac{d\epsilon}{\sqrt{2\pi}\sigma}. \tag{22}$$

Viewing the model cluster as isolated, this procedure is equivalent to averaging over those cluster states that lie within the level width above the cluster ground state.

Either of these two proceedures becomes equivalent to the classical FON method /73/ in the limit of vanishing σ. In this limit there may be convergence problems, just as in the procedure with integral occupation numbers, again for reasons of discontinuity. For values of σ significantly larger than a representative level spacing at the Fermi energy, broadening aids in SCF convergence, but for very small values of σ (compared to the gap) FON convergence can be extremely difficult to achieve. Examples may be found in many systems

(especially with rather localized or symmetry distinct orbitals near the Fermi energy), where no solution with integral occupation numbers exists that obeys 'Fermi statistics', having all occupied levels at lower one-electron energy than all unoccupied levels. In such a case, the ground state configuration is not obtained by filling orbitals from below and the lowest energy solution must be found by trial and error, just as in *ab initio* calculations.

On the other hand, FON always has a unique solution, which is guaranteed by the local potential relationship /4/,

$$\frac{\partial E}{\partial n_i} = \epsilon_i.$$

Thus the unique classical FON solution is characterized by the fact that all partially occupied orbitals have identical one-electron energies. A practical advantage gained with FON, although at some price when a solution obeying Fermi statistics exists, over using integral occupation numbers is that we never have to do more than one calculation for these difficult systems. A small σ FON SCF procedure becomes, at least temporarily, unstable if at any time during the iterations the occupation number of an isolated level near the Fermi energy changes from slightly below one to one (or from zero to a finite number). Then the Fermi energy jumps discontinuously by at least half of the difference in energy to the next higher (lower) one-electron level. However, convergence may be achieved if special precautions based on this analysis are taken /68/.

6. APPLICATION: CHEMISORPTION ON NICKEL SURFACES

Transition metal compounds present a challenge to any theoretical approach, be it *ab initio* or LDF /76,77/. The growing importance of transition metal chemistry will ensure this challenge, albeit at ever increasing system size, into the foreseeable future. Any proper description of this rich chemistry requires a balanced treatment of the valence s and d orbitals which have rather similar energies, but very different spatial character /78,79/. In conventional *ab initio* quantum chemical formulations this often forces one to go beyond HF and deal with complicated many-determinantal wavefunctions that can impede chemical insight and make accurate treatment rather difficult to achieve. The problems for an *ab initio* approach are compounded when transition metals are bonded to each other /49/. Such systems, in one sense, seem most appropriate for the approximations of LDF theory. Their high density of valence states suggests averaging and thus the statistical approach of Slater /4/. In another sense, the d electrons are too inert and localized, at least for the late first transition metal series, to strongly suggest a free-electron-gas model /5/. In any event, it is no wonder that transition metal systems have always played a prominent role among the applications of LDF methods.

The electronic structure of a large variety of transition-metal compounds has been investigated using LDF methods, particularly at fixed experimental geometries /23,25,26,30,80/. Geometry optimization is increasingly important. Full geometry optimization is possible for systems involving one transition metal using other LDF techniques /63,64/. Being able to readily optimize only a few nuclear positions in the GTO method so far, we have

focused on cluster models used for the description of local chemisorption phenomena /28,31,81,82/. In this section we discuss our LCGTO description of the chemisorption of alkali-metals and of carbon monoxide on transition metal surfaces. A proper understanding of these adsorbation systems is an indispensible prerequisite for the investigation of the phenomena observed /83/ when both these species are coadsorbed on the same surface.

Chemisorption cluster models are quite popular for studying adsorption on transition metal surfaces /28,31,82,84-88/. The cluster models the local geometry of the substrate, and usually has the spatial symmetry of the adsorption site. Cluster models can successfully give local chemical properties. Such properties include adsorption geometries, vibrational frequencies both within the adsorbate and relative to the surface, and localized excitation energies, both of core and valence electrons. Chemisorption cluster models – in their usual design – do not include lateral interactions within the adsorbate layer and are therefore most suitable for comparison with data from low coverage experiments. Our investigations of coadsorption phenomena are a step towards the more challenging problem of lateral adsorbate interactions.

In Table 1 we have collected results for atomic chemisorption from cluster calculations modelling various sites on several crystal surfaces of nickel. All the electrons are included in the SCF process. The one-electron levels were all Gaussian broadened by 0.3 eV (cf. Eq. 22). The basis sets used in these calculations have been described elsewhere /28,31,38,89,90/.

TABLE 1 - Atomic Adsorption on Nickel Surfaces. Comparison of Experimental and LCGTO-Xα results.

	Site	n[a]	h[b]		E_b[c]		ω[d]	
			calc.	exp.	calc.	exp.	calc.	exp.
Ni(111)/Na	3-fold	10	2.46	–	1.78	2.54[i]	170	–
Ni(100)/Na	1-fold	9	2.49	–	1.72	–	210	–
Ni(100)/Na	4-fold	17	2.22	2.23[e]	1.87	2.52[i]	175	–
Ni(100)/K	4-fold	17	2.67	–	1.89	–	115	–
Ni(100)/O	4-fold	9	0.95	0.86[f]	6.01	4.99[k]	425	411[l]
Ni(100)/S	4-fold	17	1.34	1.35[g]	6.23	–	320	353[m]
Ni(110)/S	4-fold	11	0.80	0.84[h]	6.18	–	385	–

[a] Number of nickel atoms in model cluster. [b] Distance of adsorbate above the first lattice plane (in Å). [c] Binding energy (in eV). [d] Vibrational frequency perpendicular to surface (in cm^{-1}). [e] /91/ [f] /92/ [g] /93/ [h] /94/ [i] /95/ [l] /96/ [m] /97/.

The results of Table 1 for the adsorption bond length and the frequency of the vibration perpendicular to the surface show good agreement with experimental data where available. The agreement between Xα and experimental binding energies is poorer, due more likely to the Xα approximation and the cluster size than to incomplete convergence in GTO basis sets. In this case the

PZ approximation would be expected to increase the calculated binding energies. The adsorption bond length of oxygen is calculated noticeably smaller than indicated by available experimental data. However, this quantity is difficult to measure in the low coverage limit to which cluster models refer because oxygen shows a strong tendency to form islands /98,99/. The ensuing inter-adsorbate repulsion and weakening of the chemisorption bond may provide an explanation for the discrepancy between experiment and cluster models in this case. It is interesting to compare our Xα results for Ni(100)/O with those of a recent large scale *ab initio* CI study based on a Ni_{41} cluster /88/, in which the following observables have been calculated: h = 1.09 Å, E_b = 4.60 eV, ω = 430 cm^{-1}.

The bonding of alkali atoms to the surface is especially interesting because they can promote catalysis. Alkali atoms are expected to easily lose their weakly bound and rather diffuse valence *s* electron, donating it to the surface. Population analysis /31/ shows the alkali substrate bonding to be fairly ionic, leading to a dipole moment at the adsorption site. The bond lengths, however, correspond to effective atomic radii for the adsorbate that are only slightly smaller than the corresponding metallic ones, but much larger than the ionic radii (e.g. for K /31/: r_{ad} = 1.95 Å, r_{ion} = 0.99 Å, r_{metal} = 2.33 Å).

TABLE 2 - Adsorption of Carbon Monoxide on Nickel. Bond lengths (in Å).

		Site	d_{Ni-CO}[a]		d_{NiC-O}[b]
		calc.	exp.	calc.	exp.
Ni(100)	2-fold	1.85	—	1.17	—
	4-fold	2.04	—	1.21	—
Ni(111)	1-fold	1.71	1.72[c]	1.16	1.15[c]
	2-fold	1.86	—	1.18	—
	3-fold	1.88	—	1.20	—

[a] Nickel carbon bond distance (in Å). [b] Carbon oxygen bond distance (in Å). [c] /100/.

In Tables 2 and 3 we present LCGTO-Xα results for the chemisorption of carbon monoxide on various sites of the surfaces Ni(100) and Ni(111). Cluster models with seven (Ni(111) two-fold) to seventeen (Ni(100) four-fold) substrate atoms have been employed. Agreement with available experimental data for the length of the adsorption and the intra-adsorbate bond and the corresponding vibrational frequencies is quite good. The lowering of the CO frequency with increasing coordination at the adsorption site is very well reproduced. We also note a concomitant lengthening of the CO bond and the distance to the nearest substrate atom. The binding energy of CO on nickel surfaces is quite small ($E_b \approx$ 1.3 eV /103/) and overestimated strongly in LDF calculations. We do not feel, therefore, that these model clusters treated at the present level of sophistication are suitable for determining preferred bonding sites.

TABLE 3 - Adsorption of Carbon Monoxide on Nickel. Vibrational Frequencies (in cm^{-1}).

	Site	$\omega_{Ni-CO}{}^a$		$\omega_{NiC-O}{}^b$	
		calc.	exp.	calc.	exp.
Ni(100)	2-fold	345	365[c]	1860	1900[c]
	4-fold	250	—	1800	—
Ni(111)	1-fold	460	—	2000	2050[d]
	2-fold	400	380[c]	1840	1831[d]
	3-fold	315	—	1780	1816[d]

[a] Vibrational frequency of adsorbate perpendicular to surface in cm^{-1}. [b] Carbon oxygen vibrational frequency in cm^{-1}. [c] /101/ [d] /102/.

The bonding of CO to transition metal atoms has been the subject of many investigations. Almost any discussion begins with the Blyholder picture /104/ of 5σ donation and $2\pi^*$ backdonation. The occupation of the antibonding $2\pi^*$ orbital of CO is held reponsible for a weakening and lengthening of the intramolecular bond, the simplest experimental diagnostic of which is a reduction in the frequency of the IR band due to C-O vibration. For gas phase CO, the bond distance, d_{CO}, is 1.128 Å and the vibrational frequency, ω_{CO}, is 2143 cm^{-1} /105/. Being more quantitative, the experimental relationship between the CO bond length (in Å) and the corresponding vibrational frequency (in cm^{-1}) has been fitted to a linear form /106/,

$$d_{CO} = -2 \times 10^{-4}\,\omega_{CO} + 1.56.$$

This relation has been confirmed (with much less scatter than found in the experimental data) by taking the Blyholder model quite literally in a series of LCGTO Xα calculations for CO /107/ with fixed fractional occupations of the 5σ and $2\pi^*$ orbitals. In this study, the change in occupation number of the 5σ orbital relative to its gas-phase value, $\Delta_{5\sigma}$, ranged between zero and -1.2, independent of $\Delta_{2\pi^*}$, which ranged from zero to 1.8. The calculated reduction of the CO frequency, $\Delta\omega_{CO}$, is fitted by the second order polynomial in the change of the occupation numbers /107/,

$$\Delta\omega_{CO} = -483\,\Delta_{2\pi^*} - 78\,\Delta_{5\sigma} - 25\,\Delta_{2\pi^*}^2 - 10\,\Delta_{2\pi^*}\,\Delta n_{5\sigma}, -66\,\Delta_{5\sigma}^2 \qquad (23)$$

to better than 10 cm^{-1} over this whole range. From the relative size of the first coefficient, the occupation of the antibonding $2\pi^*$ orbital almost exclusively determines the vibrational frequency. Chemisorption induced reduction of the CO frequency may range up to about 350 cm^{-1} on transition metal surfaces. From our fit, this would correlate with an effective $2\pi^*$ population of about 0.6 to 0.7, depending on the effective charge of the adsorbed molecule. (The bond distance is also strongly correlated with the occupation of this orbital in

this model, as it must be, based on the simple chemical concept of bonding and antibonding orbitals.)

Coadsorption with potassium on transition metal surfaces can result in dramatic lowering of the CO frequency to between 1400 and 1500 cm^{-1} /83/; for Ni(111) a frequency of 1405 cm^{-1} /108/ is observed. Cluster studies /107/ on $Ni_8(CO)K_2$ and on Ni_2CO with point charges modelling K^+ ions and their image charges (the latter describing the response of the metallic substrate) show that reductions of this size can indeed be obtained in model calculations. Beyond the Blyholder model, part of the frequency lowering induced by the coadsorbed alkali metal is due to dipole-dipole interaction /83/ between CO, on the one hand, and the adsorbed alkali ion plus its image charge, on the other hand. The charge distribution due to the coardsorbate produces a dipole layer and sets up an electric field that polarizes the CO bond. The direction of the latter is such that the bond is lengthened. (The fact that the CO IR band is so strong means that there is a very large change in its dipole moment when the CO bond distance is changed.) The indirect charge transfer from the coadsorbed potassium into the CO $2\pi^*$ orbital via the substrate and the dipole-dipole interaction contribute about equally to the lowering, both being responsible for a shift of 150 to 200 cm^{-1} in the model cluster calculations /107/. The backdonation part of this frequency shift translates into an effective population of the CO $2\pi^*$ orbital of about 1.0 to 1.1 according to Eq. 23. It is worth noting that essentially no direct orbital interaction /83/ between CO and K^+ was found in these cluster investigations /107/.

7. CONCLUSIONS

We have reviewed the present state of the Gaussian-type orbitals approach to DFT with emphasis on problems connected to the practical solution of the Kohn-Sham equations. Leaving aside questions related to the derivation of the set of one-electron equations, we have described the advances that have been made and the problems that remain when a set of one-electron equations governed by one effective local potential is to be solved for large finite systems using a GTO basis set. Given that improvements are possible on $X\alpha$ and other LDF methods /19/, we have tried to outline characteristics of local potentials that affect GTO implementation. For example, local potentials that are a function of the gradient of the charge density in general would have to be treated numerically, which in contrast to LDF methods would involve two fitting algorithms, for V and U, not just two separate fits. The gradients of the LCGTO density themselves, however, are trivially computed.

Three problem areas have been identified that plague efficient implementation of the GTO approach: the treatment of the electronic Coulomb interaction and of XC interaction as well as difficulties converging to a self-consistent solution. Variational fitting now provides a stable and economic way around the N^4 bottleneck for the electronic Coulomb interaction, reducing the problem to N^3 in the number of electrons. For the XC potential a comparable formalism is yet to be found. Available general solutions to this second problem rely on superposed atomic grids. This makes any solution costly and puts a heavy burden when accurate solutions are demanded, because accuracy, in general,

increases as the square root of the number of points. As an alternative, a stationarity principle has been discussed that is restricted to the Xα form of the XC functional, however. This latter form would have the added benefit of allowing the calculation of completely analytical energy gradients, an indispensible tool for geometry studies of complex compounds, with their often soft motions on potential surfaces embedded in high-dimensional spaces.

We have presented an application to an important chemisorption problem that also demonstrates the class of problems that may be tackled with the GTO formalism today. Not only can we study relatively simple effects in adsorbation, such as the vibrational frequencies of various adsorbed atoms and diatomics in the low coverage limit, but, given sufficient insight, we have the computational means available to attack problems in more complex adsorbate systems, such as to quantify how alkali metal coadsorption lowers the vibrational frequency of carbon monoxide so much.

ACKNOWLEDGMENT

It is a pleasure to acknowledge many stimulating discussions with E. Bertel, M. Cook, D. Menzel, J.W. Mintmire, and H.-P. Steinrück. We thank A. Görling, P. Knappe, J. Lauber and P. Sandl for their dedicated work in developing and applying the LCGTO-LDF program package. The work of N.R. was supported by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 128 and by the Fonds der Chemischen Industrie. Computer time of the Munich group was generously supplied by the Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, München.

REFERENCES

1. P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
2. N.H. March, in *Theory of the Inhomogeneous Electron Gas*, eds. S. Lundqvist and N.H. March (Plenum, New York, 1983) p. 1.
3. J.C. Slater, *Phys. Rev.* **81**, 385 (1951).
4. J.C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 4 (McGraw-Hill, New York, 1974).
5. W. Kohn and L.J. Sham, *Phys. Rev.* **140**, A1133 (1965).
6. M. Levy, *Phys. Rev. A* **26**, 1200 (1982).
7. E.H. Lieb, *Int. J. Quantum Chem.* **24**, 243 (1983).
8. G.B. Bachelet, E.R. Hamman and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982).
9. W. Hehre, L. Radom, P.v.R. Schleyer and J.A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley-Interscience, New York, 1986).
10. M.R. Pederson, R.A. Heaton and C.C. Lin, *J. Chem. Phys.* **82**, 2688 (1985).
11. P.S. Bagus and B.I. Bennett, *Int. J. Quantum Chem.* **9**, 143 (1975).
12. O. Gunnarsson and B.I. Lundqvist, *Phys. Rev. B* **13**, 4274 (1976).
13. P.-O. Löwdin, *Phys. Rev.* **9**, 1474 (1955).
14. J.W.D. Connolly, in *Modern Theoretical Chemistry*, Vol. 7, ed. G.A. Segal (Plenum, New York, 1977) p. 105.
15. B.I. Dunlap, *Adv. Chem. Phys.* **69**, 287 (1987).

16. B.I. Dunlap, *Chem. Phys.* **125**, 89 (1988).
17. B.I. Dunlap, J.W.D. Connolly and J.R. Sabin, *J. Chem. Phys.* **71**, 3396; 4993 (1979).
18. A.D. Becke, *J. Chem. Phys.* **76**, 6037 (1983).
19. A.D. Becke, *J. Chem. Phys.* **84**, 4524 (1986).
20. L. Laaksonen, D. Sundholm and P. Pyykkö, *Int. J. Quantum Chem.* **27**, 601 (1985).
21. L. Laaksonen, D. Sundholm and P. Pyykkö, *Comp. Phys. Rept.* **4**, 313 (1986).
22. D. Heinemann, B. Fricke and D. Kolb, *Chem. Phys. Lett.* **145**, 125 (1988).
23. A.D. Becke, *J. Chem. Phys.* **78**, 4787 (1983).
24. N. Rösch and H. Jörg, *J. Chem. Phys.* **84**, 5967 (1986).
25. N. Rösch, H. Jörg and B.I. Dunlap, in *Quantum Chemistry: The Challenge of Transition Metals and Coordination Compounds*, ed. A. Veillard, NATO ASI series C, Vol. 176 (Reidel, Dordrecht, 1986) p. 179.
26. N. Rösch, H. Jörg and M. Kotzian, *J. Chem. Phys.* **86**, 4038 (1987).
27. N. Rösch, P. Knappe, B.I. Dunlap, E. Bertel and F.P. Netzer, *J. Phys. C: Solid State Phys.* **21**, 3423 (1988).
28. N. Rösch, P. Sandl, A. Görling and P. Knappe, *Int. J. Quantum Chem.* **S22**, 275 (1988).
29. B.I. Dunlap, *Int. J. Quantum Chem.* **S22**, 257 (1988).
30. P. Knappe and N. Rösch, *J. Organometal. Chem.* **359**, C5 (1989).
31. N. Rösch, P. Knappe, P. Sandl, A. Görling, and B.I. Dunlap, in *The Challenge of d and f Electrons. Theory and Computation*, eds. D.R. Salahub and M.C. Zerner, ACS Symposium Series **394** (Washington 1989) p. 180.
32. H. Sambe and R.H. Felton, *J. Chem. Phys.* **62**, 1122 (1975).
33. R.A. Friesner *J. Chem. Phys.* **86**, 3522 (1987).
34. C.W. Bauschlicher Jr., S.R. Langhoff, H. Partridge and D.P. Chong, *J. Chem. Phys.* **89**, 2985 (1988).
35. B.I. Dunlap and N. Rösch, *J. Chim. Phys. Phys-chim. Biol.* **86**, 671 (1989).
36. A. Rosén D.E. Ellis, H. Adachi and F.W. Averill, *J. Chem. Phys.* **65**, 3629 (1976).
37. H. Jörg, N. Rösch, J.R. Sabin and B.I. Dunlap, *Chem. Phys. Lett.* **114**, 529 (1985).
38. P. Sandl, Diplom thesis (Technische Universität München, 1988).
39. B.I. Dunlap, *Phys. Rev. Lett.* **51**, 546 (1983).
40. P.W. Thulstrup and J. Linderberg, *Int. J. Quantum Chem.* **S13**, 39 (1979).
41. D.W. O'Neal and J. Simons, *Int. J. Quantum Chem.*, to be published.
42. J.W. Mintmire, J.R. Sabin and S.B. Trickey, *Phys. Rev. B* **26**, 1743 (1982).
43. J.W. Mintmire and C.T. White, *Phys. Rev. B* **28**, 3283 (1983).
44. J.P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).

45. S.H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).
46. D.M. Ceperley and B.J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
47. R. Gaspar, *Acta Phys. Acad. Sci. Hung.* **3**, 263 (1954).
48. E.J. Baerends and P. Ros, *Int. J. Quantum Chem.* **S12**, 169 (1978).
49. D.R. Salahub, *Adv. Chem. Phys.* **69**, 447 (1987).
50. J. Harris and G.S. Painter, *Phys. Rev. B* **22**, 2614 (1980).
51. K. Lee, J. Callaway and S. Dhar, *Phys. Rev. B* **30**, 1724 (1984).
52. P.M. Boerrigter, Ph.D. dissertation (Free University of Amsterdam, 1987).
53. B.I. Dunlap and M. Cook, *Int. J. Quantum Chem.* **29**, 767 (1986).
54. J.L. Martins, R. Car and J. Buttet, *J. Chem. Phys.* **78**, 5646 (1983).
55. C.F. Fischer, *The Hartree-Fock Method for Atoms* (Wiley, New York, 1977).
56. F. Herman and S. Skillman, *Atomic Structure Calculations* (Prentice-Hall, Englewood Cliffs, 1963).
57. A.H. Stroud, *Approximate Calculations of Multiple Integrals* (Prentice-Hall, Englewood Cliffs, 1971).
58. R.S. Jones, J.W. Mintmire and B.I. Dunlap, *Int. J. Quantum Chem.* **S22**, 77 (1988).
59. E.J. Baerends, D.E. Ellis and P. Ros, *Chem. Phys.* **2**, 41 (1973).
60. B.I. Dunlap and W.N. Mei, *J. Chem. Phys.* **78**, 4997 (1983).
61. J.W. Mintmire, unpublished.
62. T. Ziegler and A. Rauk, *Theor. Chim. Acta* **46**, 1 (1977).
63. L. Versluis and T. Ziegler, *J. Chem.Phys.* **88**, 322 (1988).
64. L. Versluis, Ph.D. thesis (University of Calgary, 1989).
65. B.I. Dunlap, *J. Phys. Chem.* **90**, 5524 (1986).
66. J.W. Mintmire, *Int. J. Quantum Chem.* **S13**, 165 (1979).
67. J.W. Mintmire and J.R. Sabin, *Chem. Phys.* **50**, 91 (1980).
68. B.I. Dunlap, unpublished.
69. B.I. Dunlap, *Phys. Rev. A* **25**, 2847 (1982).
70. J.H. Wood and M. Boring, *Comp. Phys. Comm.* **7**, 73 (1974).
71. M.C. Zerner and M. Hehenberger, *Chem. Phys. Lett.* **62**, 550 (1979).
72. N. Rösch and P. Knappe, unpublished.
73. J.C. Slater, J.B. Mann, T.M. Wilson and J.H. Wood, *Phys. Rev.,* **184**, 672 (1969).
74. S. Dahr, *Phys. Rev. A* **39**, 952 (1989).
75. B.I. Dunlap, *Phys. Rev. A* **29**, 2902 (1984).
76. A. Veillard, ed., *Quantum Chemistry: The Challenge of Transition Metals and Coordination Compounds*, NATO ASI series C, Vol. 176 (Reidel, Dordrecht, 1986).
77. D.R. Salahub and M.C. Zerner, eds., *The Challenge of d and f Electrons. Theory and Computation*, ACS Symposium Series **394** (Washington 1989) p. 180.
78. C.W. Bauschlicher Jr. and P.S. Bagus, *J. Chem. Phys.* **91**, 5889 (1984).
79. H. Jörg and N. Rösch, *Chem. Phys. Lett.* **120**, 359 (1985).

80. D.A. Case, *Ann. Rev. Phys. Chem.* **33**, 151 (1982).
81. H. Jörg and N. Rösch, *Surface Sci.* **163**, L627 (1985).
82. N. Rösch, P. Sandl, P. Knappe, A. Goerling, B.I. Dunlap, *Z. Phys. D* **12**, 547 (1989).
83. H.P. Bonzel, *Surface Sci. Rep.* **8**, 43 (1987).
84. D. Post and E.J. Baerends, *J. Chem. Phys.* **78**, 5663 (1983).
85. J. Andzelm and D.R. Salahub, *Int. J. Quantum Chem.* **29**, 1091 (1986).
86. B. Wästberg and A. Rosén, *Surface Sci.* **193**, L7 (1988).
87. G. Pacchioni and J. Koutecký, in ref. 77.
88. I. Panas, P. Siegbahn and U. Wahlgren, *Theor. Chim. Acta* **74**, 167 (1988).
89. F.J. van Duijneveldt, *IBM Research Report*, No. RJ 945 (1971).
90. A. Veillard, *Theor. Chim. Acta* **12**, 405 (1968).
91. A. Winkler, *J. Phys. C: Solid State Physics* **16**, L99 (1983).
92. A. Winkler, K.D. Rendulic and K. Wendl, *Appl. Surf. Sci.* **14**, 209 (1982).
93. J.E. Demuth, D.W. Jepsen and P.M. Marcus, *J. Phys. C: Solid State Physics* **8**, L25 (1975).
94. J. Stoehr, R. Jaeger and T. Kendelewicz, *Phys. Rev. Lett.* **49**, 142 (1982).
95. Z.Q. Wu, Y. Chen, M.L. Xu, S.Y. Tong, S. Lehwald and H. Ibach, *Phys. Rev. B* **39**, 3116 (1989).
96. R. Baudoing, Y. Gauthier and Y. Joly, *J. Phys. C: Solid State Physics* **18**, 4061 (1985).
97. R.L. Gerlach and T.N.Rhodin, *Surface Sci.* **17**, 32 (1969).
98. D. Brennan, D.O. Hayward and B.M.W. Trapne, *Proc. Roy. Soc. A* **256**, 81 (1960).
99. R.Franchy, M.Wuttig and H.Ibach, *Surface Sci.* **203**, 489 (1988).
100. M. Passler, A. Ignatiev, F. Jona, D.W. Jepsen and P.M. Marcus, *Phys. Rev. Lett.* **43**, 360 (1979).
101. J.C. Bertolini and B. Tardy, *Surface Sci.* **102**, 131 (1981).
102. L. Surnev, Z. Xu and J.T. Yates Jr., *Surface Sci.* **201**, 14 (1988).
103. K. Christmann, O. Schober and G. Ertl, *J. Chem. Phys.* **60**, 4179 (1974).
104. G. Blyholder, *J. Phys. Chem.* **68**, 2772 (1964).
105. K.P. Huber and G. Herzberg *Molecular Spectra and Molecular Structure*, Vol. 4: *Constants of Diatomic Molecules* (Van Nostrand, Princeton, 1979) p. 166.
106. D.F. Ogletree, M.A. Van Hove and G.A. Somorjai, *Surface Sci.* **173**, 351 (1986).
107. N. Rösch, J. Lauber, A. Görling and P. Knappe, to be published.
108. L. Ng, J. Uram, Z. Xu, P.L. Jones and J.T. Yates Jr., *J. Chem. Phys.* **86**, 6523 (1987).

LOCAL DENSITY FUNCTIONAL THEORIES OF IONIC AND MOLECULAR SOLIDS

Roy G. Gordon

Department of Chemistry
Harvard University
Cambridge, Mass., 02138

Richard LeSar

Los Alamos National Laboratories
Los Alamos, New Mexico 87545

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

1. INTRODUCTION

The pioneering paper by Hohenberg and Kohn[1] 25 years ago put density-functional methods for electronic structure calculations on a firm theoretical footing. Techniques such as those developed by Kohn and Sham[2] are mainstays in calculating properties of solids, especially metallic systems, and are discussed in detail in other articles in this volume. Here we will describe a rather different application of density-functional theory, an approximate method called the (Gordon-Kim) electron-gas model[3], that is useful for calculating interaction energies between closed-shell atoms and molecules.

Fundamental to almost all discussions of molecular systems is the assumption that these systems consist of molecular units, with the electrons tightly bound to a specific molecule. Molecular and ionic solids, for instance, are characterized by having electrons highly localized about their parent nuclei, with very little electronic density in the interstitial regions. This extreme electron localization allows a great simplification in the application of local-density functional theories. To a first approximation, the electronic structure of interacting molecules can be taken as a sum of the unperturbed (i.e. gas-phase) electron densities of those molecules. The evaluation of the energy is then simple; the electronic structure is a given and the interaction energy alone needs to be evaluated (i.e. no self energies). Since calculation of the electronic structure of isolated molecules is straightforward (for example with Hartree-Fock theory), the input densities are accurately known and easily obtainable. Also, since the electronic structure is input, and since the interaction energies are dominated by the electrons in the region between the interacting molecules and not the core electrons, the kinetic energy operator can be replaced with the free-electron gas kinetic energy density functional without great loss of accuracy. Because of its simplicity, and the rather accurate results obtained with the model, the electron-gas model is one the most useful methods for

determining interaction energies between closed-shell atoms, molecules and ions.

It turns out that the assumption of gas-phase electronic densities works well for pair interactions, but is not so good for some solids, particularly ionic solids, where the large electrostatic interactions in the crystal can considerably modify the electronic structure of the ions. Much of the rest of this article is concerned with the problem of how to best approximate the electronic density of atoms and molecules in solids. It should be noted that in all the electron-gas-model calculations discussed here, the total electron density is still represented as a sum of the distributions of the constituent molecules. However, instead of using gas-phase molecular charge distributions, one uses charge densities perturbed by the crystalline environment around the molecules.

2. THE ELECTRON-GAS MODEL

In its original form,[3] the electron gas model calculates the short-range interaction energy between closed-shell systems via two simple approximations. First, the total electronic density of a combined system is assumed to be a sum of the unperturbed (i.e. gas-phase) densities of the constituent systems (the additive density approximation);

$$\rho(\mathbf{r}) = \sum_i \rho_i(\mathbf{r}) \quad , \qquad (1)$$

where ρ_i is the electronic charge density of constituent i. Second, the energy density is approximated by the energy density of a homogeneous electron gas, $e[\rho]$, with the local density approximation, i.e. the energy density at a point depends only on the charge density at that point. The interaction energy is then obtained by numerically integrating the difference between the energy density of the combined system and the sum of the energy densities of the constituent systems. For a pair of interacting systems located at $\mathbf{r}_a$ and $\mathbf{r}_b$, we have

$$E_{ab} = \int dr \left\{ e\left[\rho_a(r - r_a) + \rho_b(r - r_b)\right] - e\left[\rho_a(r - r_a)\right] - e\left[\rho_b(r - r_b)\right] \right\} \quad (2)$$

where

$$e[\rho] = C_k \rho^{5/3} - C_x \rho^{4/3} + e_c[\rho] + \rho_T \Phi[\rho_T] \quad (3)$$

and the first three terms represent the kinetic, exchange, and correlational energy densities, respectively. The constants $C_k = 3/10(3\pi^2)^{2/3}$ and $C_x = 3/4(3/\pi)^{1/3}$ are the standard constants for a uniform electron gas.[4] The form of E_c was determined by extrapolating between the known large and small density behavior and is given elsewhere.[3] The last term is the Coulomb energy, where ρ_T is the total charge density (nuclear and electronic) and Φ is the Coulomb potential. Despite the approximations inherent in both the description of charge density and the energy density functionals, this simple model is remarkably successful at predicting the short-range repulsive part of the intermolecular potential for pairs of closed-shell atoms, molecules and ions. Early applications and approximations of the electron-gas model have been reviewed previously.[5]

A number of approaches have been suggested to modify and improve the energy-density functionals used in Eq. 2. Perhaps the most used method is the modified electron-gas (MEG) model.[6] By comparing total kinetic and exchange energies of atoms calculated both with the electron-gas terms of Eq. 2 and with Hartree-Fock calculations, Waldman and Gordon[6] found that the electron-gas results could be brought close to agreement with the Hartree-Fock calculations by scaling the electron-gas terms with coefficients that depended only on the number of electrons. These same scaling factors can then be used to multiply the electron-gas terms for use in calculating interaction energies, with the number of electrons now for the combined system. Use of the MEG approximation leads to improved agreement between experimental and calculated interaction potentials.[6] Because the original form of the local-density correlation functional was an

approximate representation, some groups[7] have used the presumedly more accurate Hedin-Lundqvist [8] parametrization. Finally, there has been some study of non-local-density approximations in the context of the electron-gas model. The work as applied to the Gordon-Kim model is reviewed by Pearson and Gordon[9] in a paper on the use of gradient corrections to the uniform electron-gas functionals. They show that proper inclusion of gradient corrections (by treating the expansion as locally asymptotic) to the kinetic energy may improve the results obtained with the electron-gas model.

Long-range dispersion (van der Waals) interactions, which arise from the interaction of instantaneous dipoles created by the fluctuating electronic distributions, giving the well-known r^{-6} form at long range, and are not accounted for in the electron-gas model. At small separations, however, these interactions are part of the correlational energy. A number of approximate ways to include these interactions have been studied, including the use of a Drude oscillator model;[10,11] matching the known long-range form onto the correlational energy at small separations[12], and semiempirical approaches.[13,14,15] Because of ambiguities about how to match up the long-range and short-range terms, none of the suggested methods have been entirely satisfactory.

In addition to the approximations inherent in the use of local-density functionals, the electron-gas model makes assumptions about the electronic density of the interacting systems. The principal assumption is the additive density approximation, where the total electronic density is taken as the sum of the gas-phase distributions of the interacting molecules. For rare-gas interactions, under normal conditions, this seems to be a reasonable assumption. Calculations by Wood and Pyper[16] on these systems suggest that by not relaxing the charge distributions, the repulsive wall is shifted to somewhat larger separations, but that the errors are not large. As noted in the introduction, however, in ionic solids the charge distributions of the constituent atoms are perturbed from the gas phase. In the next section, we

discuss approximate ways to calculate the perturbed electronic density of atoms, ions and molecules inside solids.

In condensed phases, there is always the question of whether the total interaction energy can be represented as a sum of interactions between pairs of atoms. To test these effects, Gordon and co-workers[17,18] introduced methods by which the many-body, short-range interaction energy could be determined directly by a numerical integration of the energy-density functionals over the entire unit cell. In ionic solids, these many-body energies are reasonably small, as we shall see in comparisons below.

2.1. Free ion approximation (FIEG and FIMEG)

The earliest applications of the electron gas (EG) model to ionic solids[19] assumed simply that free, gas-phase-phase ions were placed unchanged in the crystal (FIEG). Electron densities for the free ions were evaluated from atomic Hartree-Fock self-consistent-field wave functions. The interaction potentials of the nearest neighbor pairs of ions were determined as in Eq. 2. The electrostatic interactions between all the ions in the crystal were evaluated using the Madelung constant for the appropriate crystal structure.

The results with this very simple approach were reasonably accurate for the alkali halide and alkaline earth dihalide crystals. Lattice constants and lattice energies for the alkali halides were found to be within about 4% and 8% of experiment, respectively. The bulk moduli, which involve second derivatives of the potential, were less well determined, underestimating the experimental values by up to 30%.

Applications of the free ion approximation were also made with the MEG model (FIMEG), both with pair interactions[20] and including the many-body short-range energy.[17] Results with the pair interactions were surprisingly accurate for the alkali halide crystals, with the lattice parameters and energies within about 2% of experiment. The calculated results for the bulk moduli were improved to within about 10% of the experimental values. Inclusion of the many-body energy made very little change in the calculated

properties.[17] Phase transitions between different crystal structures were also predicted,[20] but the pressures of these phase transitions showed large errors because of the extremely small energy differences involved.

The free-ion approximation is limited in its range of applicability to ions which are stable in the gas phase. All positive ions are stable as free ions, so the free-ion approximation can be used for any cations. However, only anions with a single negative charge are stable, such as halide (F^-, Cl^-, Br^-, I^-), hydride (H^-), hydroxide (OH^-) and cyanide (CN^-) ions. Multiply charged anions such as oxide (O^{2-}) are not stable in the gas phase, and thus the free-ion approximation cannot be applied to oxide crystals.

2.2. Spherically stabilized rigid ions (RIMEG)

Since oxide ions do not exist in free space, one may question whether they exist in oxide crystals. Generations of elementary chemistry students have been taught about oxide ions as constituents of ionic compounds. Have we been misleading them? We think not, since the oxide ion is stabilized by its crystalline environment, which must carry two net positive charges to balance the charge of the oxide ion. These positive charges around the oxide ion help bind two electrons to an oxygen atom in a crystal to form a crystal-field stabilized oxide ion.

A simple model for this crystal-field stabilization was introduced by Watson,[21] who distributed the positive charge on a spherical shell surrounding the oxide ion. The additional potential produced by this charge shell is constant within the shell, and Coulombic outside the shell. The Hartree-Fock wavefunction for such a stabilized oxide ion retains the spherical atomic symmetry, and can thus be obtained with simple modifications of programs for atomic wave functions.

The electron densities for the Watson-sphere stabilized oxide ion were used by Cohen and Gordon[22] to calculate energies for magnesium and

calcium oxide crystals. The resulting lattice constants and energies were in reasonable agreement with experiment.

In making comparison with the lattice energy, care must be taken to use a consistent and well-defined reference state, since there is some confusion in the literature on this point. In particular, attempts to use "gas-phase O^{2-} ions" as a thermodynamic state are meaningless and undefined because of the instability of this species. One consistent and well-defined gaseous reference state consists of positive ions, singly-charged (stable) O^- ions and free electrons. One of the energy terms needed in assembling the lattice is the energy required to add a free electron to an O^- ion, to form a Watson-sphere stabilized O^{2-} ion. This energy, which we call the self-energy of the O^{2-} ion, is added to the Coulomb terms (Madelung energy) and the pair potentials (calculated from the electron-gas energy functional) to find the (static) lattice energy.

2.3. Variationally stabilized non-rigid spherical ions (VSMEG)

There remains the question of what charge and radius should be used to describe the crystal field stabilization. Watson originally used a charge of +1 electronic charge and a shell radius of 2.66 au, a value estimated as the empirical "ionic radius" of the oxide ion. While these shell parameters give reasonable results, most later workers[17] used a shell charge equal in magnitude and opposite in sign to the charge on the ion being stabilized. Thus for an O^{2-} ion, a +2 charge shell is normally used to represent the crystal field. Recently, however, Jackson and Gordon[23] have argued that a shell charge slightly less in magnitude than the ion charge may be more appropriate for crystals at high density, because of the increased overlap of ionic charge densities at high pressures. Shell charges of 1.7 to 1.8 gave slightly more stable lattices for two magnesium oxide phases, than did the conventional shell charges of 2.0.

The choice of the shell radius has been the subject of much discussion and calculation. Mulhausen and Gordon[17] originally used the shell radius as a variational parameter in order to find the the minimum crystal energy as a function of the shell radius (see figure 1), at the same time establishing the most appropriate shell radius for a given crystal. However, because the crystal energy is so flat in the neighborhood

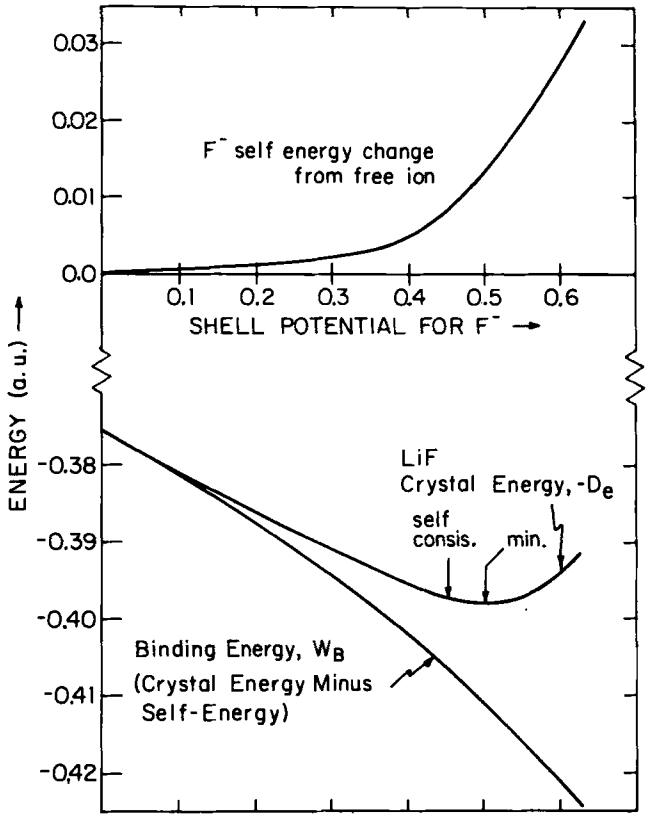


FIG. 1. Variation of self-energy, crystal energy, and binding energy with shell potential in LiF. The self-consistent and minimum crystal energy points are indicated. (from ref. 17)

of the minimum, they found it difficult to determine reliable values for the shell radius by this variational method. Therefore they used an electrostatic

consistency argument, discussed below, to estimate the shell radius in his calculations.

Recently, Wolf and Bukowinski[24] have successfully implemented the variational approach to determining shell radii. They found that larger basis sets are needed for the Hartree-Fock ionic wave functions, and that very careful optimization of the orbital exponents in these wave functions is required for each shell radius. With these precautions, they were able to calculate very reliable equations of state and other properties for magnesium and calcium oxides, using the variational approach. Even the elastic constants are predicted fairly well. It is interesting to note that the Cauchy relations (e.g. $C_{12} = C_{44}$ for these crystals) are violated by the theory (and the experiments). Normally, all theories based on spherical pair interactions incorrectly predict the Cauchy relations to hold. However, the pair potentials in the present theory depend implicitly on the rest of the lattice, so they are really many-body forces.

The variationally determined shell radii decrease as the crystal volume decreases at high pressures. The corresponding ionic electron densities also shrink in size as the pressure increases. This squeezing of the ions to smaller size is an important physical effect, particularly in predicting properties such as the bulk modulus. If the ions are considered as rigid, the predicted bulk moduli will be too large, and the lattice too rigid.

2.4. Coulomb-stabilized non-rigid spherical ions (CSMEG)

Because of the numerical instabilities in the variationally determined shell radii, Mulhausen and Gordon[17] adopted a self-consistency argument, first suggested by Pachalis and Weiss[25], to estimate the shell radii. Since the Watson sphere shell potential is supposed to represent the stabilizing effects of the charges in the surrounding lattice, the Coulomb potential at the site of the ion was calculated, and the shell radius adjusted so that the Coulomb potential inside the sphere matched the site potential. Since

changing the lattice constant changes the site potential, and hence the shell radius, the ion charge density and the pair potentials, it was necessary to determine the shell radius and the equilibrium lattice constant iteratively. However, the iterations converge rapidly, and the shell radius is determined easily without numerical instability. Equilibrium lattice constants and energies were predicted with a typical accuracy of 1-2% for a wide variety of ionic crystals. Equations of state were less reliable, in part because the pressure due to the implicit volume-dependence of the self-energy was neglected. By including this term, Coulomb-stabilized non-rigid spherical ionic calculations of equations of state were successfully predicted for NaF and NaCl[26] and MgO.[27] The computer program used for finding these equations of state was specific to these symmetric crystals without internal parameters.

This same model was applied to less symmetric crystals and to their harmonic vibrations by Boyer et al.[28] They renamed it the potential-induced breathing (PIB) model, and were apparently unaware that it was identical to the Coulomb-stabilized version of the model used by Mulhausen and Gordon.[17] Boyer et al.[28] use a different local density functional for the pair interactions, and also use it, rather than Hartree-Fock, to determine ion wave functions and self-energies. These differences, not their claimed differences in stabilization method, which account for the difference between CSMEG and PIB results for the structure and energy of MgO and CaO. Elastic constants and lattice vibration frequencies calculated by Boyer et al.[27] showed improved agreement with experiment, compared to rigid ion models.

The shell radii determined by this Coulomb consistency argument are close to, but not identical to, those determined from the variational argument (see figure 1 of reference 17). The differences between crystal properties predicted by the variational and Coulomb-stabilized methods are small, and probably within the accuracy of the modified electron-gas density functional (see Wolf and Bukowinski,[24] for a good comparison of these results).

2.5. Non-spherical, polarized ions (VPMEG)

When ions are in a non-centrosymmetric environment, their electron densities are distorted or polarized. An extreme example is an ion in a diatomic molecule, such as an alkali halide molecule. An extension of the electron-gas model was developed[10] to include such polarization effects. The electron density in the outermost (valence) orbitals is assumed to shift rigidly under the influence of the following forces on the valence electrons: 1. Coulomb interactions due to the other ion(s) in the system. 2. Forces due to overlap of the valence electrons with electrons on other ion(s) in the system, computed in the usual way by integrating an energy density functional over the region of overlap. 3. A restoring force between the valence electron and the nucleus and core electrons of the ion to which the valence shell belongs. This restoring force is assumed to be harmonic, and the force constant is determined by the dipole polarizability of the ion, determined either theoretically or experimentally. The equilibrium position of the valence shell under these forces is determined. Thus this method is a close relative to the VSMEG method, except that the charge relaxation is a shift rather than a breathing motion.

This method was applied successfully to many pair interactions of atoms and ions[10] and molecules.[11] Application of a similar model to certain solids, such as alumina, gave promising results.[29] This model is currently being applied to other non-centrosymmetric solids and to lattice vibrations which distort even centrosymmetric solids away from a symmetric ionic environment.[30]

Application of this single-shell polarization model to other solids, such as silica, did not give meaningful results. Because the oxide ions in silica are between two equally close silicon neighbors, there is no preference for shifting the shell toward either of these neighbors, and the shell ends up shifting in a direction halfway between the silicon neighbors. Therefore a two-shell polarization model was devised for this kind of situation.[31] The

valence electron density is divided into two equal shells, each of which is allowed to shift independently. In addition to the forces enumerated above for the single-shell model, there is a force between the shells, which is estimated from the theoretical quadrupolar polarizability of the oxide ion. This two-shell polarization model predicts the energy, structure and compressibility of alpha- and beta-quartz[30] and coesite[32] much better than the unpolarized MEG methods do. In particular, the two-shell model correctly finds that alpha-quartz is more stable than beta-quartz, whereas unpolarized models incorrectly find the reverse.[33] The silicon-oxygen-silicon bond angles are nearly correct in the two-shell model, whereas unpolarized models find larger, nearly linear bond angles. The bulk modulus, which is much too large in unpolarized models, is reduced to reasonable values by the two-shell model, which correctly accommodates the compression by bending the bond angles. Application of the two-shell model to olivine and spinel forms of Mg_2SiO_4 resulted in too much polarization;[34] since there are four cations close to an oxygen in this case, it may be necessary to use a more elaborate polarization model.

2.6. Ionic crystals containing polyatomic anions

The application of these ideas to systems involving molecular ions is complicated by the non-spherical nature of the molecular charge distributions. Most of the work done on molecular ionic systems has been in the spirit of the CSMEG model. In studies of the alkali and alkaline-earth hydroxides[18] and the alkali cyanides,[35] the lattice-Coulomb potential was calculated at a series of sites along the OH^- and CN^- bonds. This Coulomb potential was then approximated with a set of six bare charges (replacing the Watson sphere) and a Hartree-Fock calculation was performed on the ion in the presence of these charges. As in the CSMEG approach, the calculations were repeated until self-consistency between the calculated site potentials and those used in determining the wave functions was reached. The effect of the ionic environment on the molecular charge distributions was similar

to the that found for atomic ions, with a decrease in calculated lattice constants when the stabilized distributions were used. The agreement with experiment of these calculations was good; within a few percent in equilibrium lattice energies and constants. Analysis of the molecular wave functions showed a modest change in the electrostatic moments of the molecules, though the use of rather small basis sets in the Hartee-Fock calculations (*s* and *p* orbitals only) limited the accuracy of these calculations. As an example, the dipole moment of a CN^- ion was found to be reduced from about 0.19 atomic units in the gas-phase to 0.09 in low-temperature solid NaCN, with almost no change in the higher moments.[35] Recently, Fowler and Klein[36] did a more careful determination of electronic structure changes in CN^- ions in the high-temperature rotor phase of NaCN. They approximated the crystal environment by fixing a CN^- ion and its six nearest neighbor Na^+ ions in their known lattice sites and performing Hartree-Fock calculations on this cluster, thus neglecting long-range contributions from the rest of the lattice. However, using a more complete basis set that included *d* functions, they found a larger reduction in the dipole moment (from about 0.15 to 0.03 atomic units) with a reasonably large reduction (about 20%) in the quadrupole moment as well.

2.7. Neutral molecular crystal perturbation model (CPM)

Up to now, we have discussed approximate ways of including the effects of electrostatic interactions on the electronic structure of atoms and molecules in ionic solids. For systems made up of neutral species, for instance the rare-gas solids or solid nitrogen, the electrostatic interactions between molecules are reasonably small and do not lead to large perturbations of the molecular electronic structures. At high pressures, however, there is significant overlap of the molecular charge distributions, leading to changes in the electronic structure from the orthogonality of the wave functions. LeSar[37,38] introduced a model to approximate these effects that is in the spirit of the CSMEG model for ionic systems. In this

approach, the Hamiltonian for a molecule in a solid is written as a sum of the free-molecule Hamiltonian plus a term that represents the interaction of the electrons in the molecules with the other electrons in the solid. If only the electrostatic interactions were included, this approach would reduce to the CSMEG model. However, an approximate potential, derived from the energy-density functional of Eq 2, was suggested to represent the repulsive interactions of the electrons with the rest of the atoms. Application of this approach to solid Ar[37] and He[38] found that inclusion of these terms leads to a compression of the electronic structure of the atoms as the pressure is increased. If this atomic compression is included in determining the interaction between the atoms, greatly improved agreement between experimental and theoretical equations of state is obtained.

3. COMPARISON OF METHODS

Theoretical Method	% error in lattice constant	% error in dissociation energy	% error in bulk modulus
FIEG[a]	+1.4	+1.6	
FIMEG - pair[b]	+3.9	-4.1	-7±7
FIMEG - MB[c]	+5.2	-5.9	
CSMEG - MB[c]	+1.6	-3.3	
CSEG - TF[d]	-1.3	?	-7±7
CSEG - KS[d]	-2.1	?	+7±7
CSMEG[e]	+0.7	-0.7	+9±7
APW[f]	+0.1	?	0±7

[a] Kim and Gordon (1974), ref. 19
[b] Cohen and Gordon (1975), ref. 20
[c] Mulhausen and Gordon (1981), ref. 17
[d] J. L. Feldman et al, Phys. Rev. B37, 4784 (1988)
[e] Hemley and Gordon (1985), ref. 26
[f] M. S. T. Bukowinski and J. Aidun, J. Geophys. Res. 90, 1794 (1985)

Table 1. Comparison of results for NaCl (static lattice) by various electron gas density functional methods.

In Tables 1-3 we compare results of electron-gas model calculations on three systems, NaCl, MgO, and SiO_2, using the various approximations to the charge density discussed above. By comparing calculated values for the lattice constant, dissociation energy, and bulk modulus to experiment, we can examine each of the suggested descriptions.

The results for NaCl are shown in Table 1. NaCl is a case for which the electron-gas model should work well in its simplest form, i.e. with the assumption of spherical ions. NaCl is highly ionic, and X-ray determinations[39] of its electronic structure indicate that the ions are essentially spherical, with minimal electron distribution in the interstitial regions.

The first entry in Table 1 is from a calculation involving free-ion wave functions and the EG model (FIEG). The relatively small deviations from experiment in the FIEG calculation are fortuitous and most likely are due to cancellation of errors between the more approximate energy-density functionals and the gas-phase electronic density. A more realistic assessment of the use of gas-phase wave functions is given by the FIMEG-pair and FIMEG-MB calculations, which used the free-ion wave function and the more accurate modified electron-gas (MEG) model. These calculations differ from each other in the treatment of the short-range interaction energy. The pair calculations assume pairwise additive potentials while the MB calculations included contributions to the short-range MEG energy from the mutual overlap of more than two charge distributions. Both of these calculations have lattice constants that are larger and dissociation energies that are smaller than experiment. This trend is typical of results for ionic solids. Use of gas-phase charge densities yield lattices that are too large and binding energies that are too small. Inclusion of the Coulomb charge stabilization (CSMEG-MB) leads to a marked improvement in the agreement with experiment both for the lattice parameter and the dissociation energy. This calculation did not include the self-energy contribution to the pressure, which is included in the calculation marked

CSMEG. Including this term leads to very close agreement with experiment, though, since pair interaction potentials were used, it is difficult to compare the CSMEG calculations with the CSMEG-MB calculations. The other two electron-gas calculations listed the CSEG-TF and CSEG-KS used Coulomb-stabilized ions, whose wave functions were determined with either the Thomas-Fermi kinetic energy functional or the Kohn-Sham method, and the EG model with the Hedin-Lundqvist exchange-correlation term. The results with these calculations are somewhat in poorer agreement with experiment than the CSMEG method. Finally, we report the results from accurate APW (augmented-plane wave) band-structure calculations, which show the best agreement with experiment, as they should since the change in electronic structure is calculated very accurately without the errors introduced by a model. The price paid in the APW calculations is the restriction to symmetric structures and the computer time needed in the calculation.

Theoretical Method	% error in Lattice constant	% error in dissociation energy	% error in bulk modulus
RIMEG[a]	+9.2	-8.8	-2.4
CSEG[b]	+2.6	-32.6	-19.4
CSMEG[c]	+2.4	-0.7	+15.5
VSMEG[d]	+0.4	-0.2	+10.6
APW[e]	+0.3	?	-3.2

[a] Cohen and Gordon, ref. 20
[b] Mehl et al. (1986), ref. 7
[c] Hemley et al. (1985), ref. 27
[d] Wolf and Bukowinski (1988), ref. 24
[e] M. S. T. Bukowinski, Geophys. Res. Letters 12, 536 (1985)

Table 2. Comparison of results for MgO (0°K) by various electron-gas density functional methods.

In Table 2, results of calculations on MgO are compared. Overall, agreement between calculation and experiment for this system is comparable

to that for NaCl. The CSEG results used the standard Thomas-Fermi kinetic energy functional with the Hedin-Lundqvist exchange-correlation term. As can be seen in the table, use of this approach does not yield as close agreement with experiment for MgO as calculations based on the MEG model, which uses the Thomas-Fermi-Dirac functional with each term scaled by a factor that depends on the number of electrons in the system as described above. The best agreement of the electron-gas models with experiment is with the VSMEG, which treats the radius of the Watson sphere as a variational parameter. The VSMEG method for MgO works about as well as the accurate APW results. The RIMEG calculations used Watson's wave function for the O^{2-} ions and show rather poor agreement with experiment. This disagreement can be understood by examining the oxide wave function. Comparison of $\langle r^2 \rangle$ of the electronic distribution shows that the Watson wave function has a considerably larger extent than the converged wave functions determined in the CSMEG method, with the value of the Watson wave function about 27.53 $bohr^2$ and CSMEG result of 23.00 $bohr^2$. The larger extent of the wave function leads to a larger lattice than experiment and a smaller dissociation energy.

Theoretical Method	% error in (lattice volume)$^{1/3}$	% error in dissociation energy	% error in bulk modulus
CSMEG[a]	+0.9	-5.0	+710
VPMEG[a]	-1.3	+1.0	+49

[a] Jackson and Gordon, ref 31

Table 3. Comparison of results for SiO_2 in α-quartz structure (static lattice) by electron gas density functional methods.

The final example shows the importance of polarization for non-centrosymmetric solids. In SiO_2, the oxide ions are polarized toward their two silicon nearest neighbors. As discussed above and shown in Table 3, use of spherical non-polarized ions (CSMEG) leads to results for the

structure of the solid in much worse agreement with experiment than the use of the two-shell polarized model (VPMEG). In particular, the silicon-oxygen-silicon bond angle is well determined with polarization and not without polarization (Fig.2).

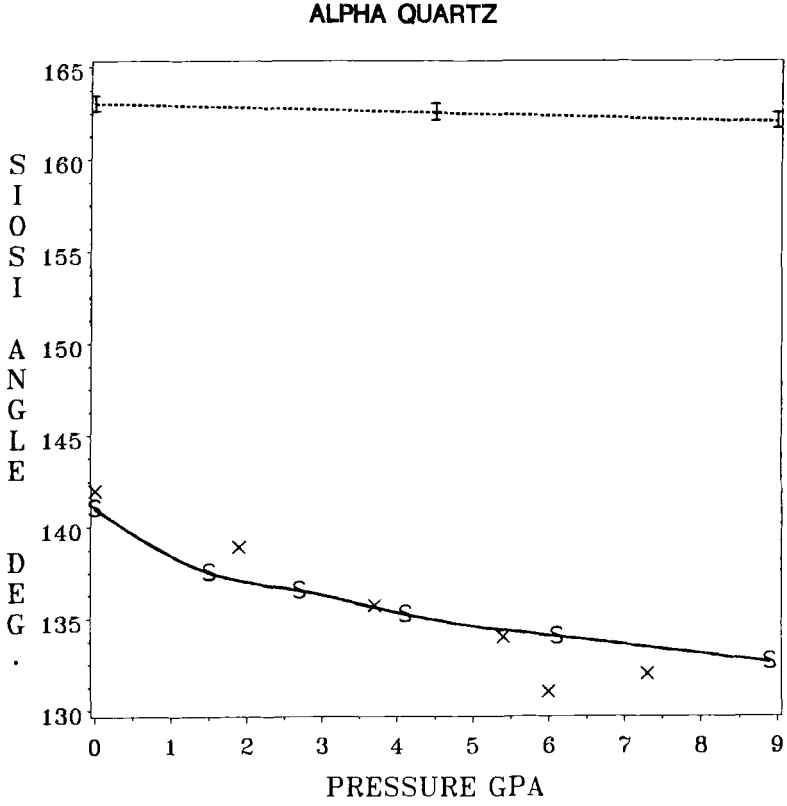


FIG. 2. Pressure variation of the Si-O-Si angle in α quartz: Shell model -S-, Experiment X, and Ionic model -I-

The calculated bulk modulus with the CSMEG model is too high by a factor of 7, and with the two-shell polarization is reduced to reasonable agreement with experiment. Considering the complexities of the quartz structure, the simple two-shell model works remarkably well.

Much has been said in this article about the role of the solid-state environment in determining the electronic structure of the constituent ions

and thereby the interactions between the ions, which in turn determine the structure. A sensitive indicator of the changes in the electronic distributions is the $\langle r^2 \rangle$ of the electronic distribution. In Figure 3 we plot the ratio of $\langle r^2 \rangle$ in the solid to that in the gas phase, $\langle r^2 \rangle_g$, for a series of ions involving F^- and Cl^- (diamonds) anions.

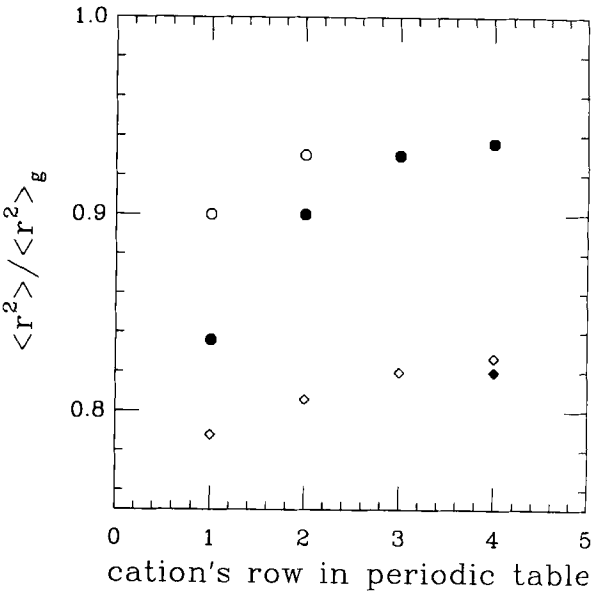


Fig. 3. Ratio of the mean-square radius of the crystal-stabilized wave function. ($\langle r^2 \rangle$) to the gas-phase value. ($\langle r^2 \rangle_g$) for Cl^- (diamonds) and F^- (circles) ions in a series of ionic solids. The row in the periodic table of the cation in the solid is given as the obscissa. Open symbols are for monovalent cations (e.g. NaF) and closed symbols are for divalent cations (e.g. NgF_2).

This ratio is plotted versus the row of the periodic table of the cation in the ionic solid, with the open symbols for single-valent cations (eg. Na^+) and the closed symbols for divalent cations (eg. Mg^{2+}). This ratio varies for F^- ions from about 0.9 to 0.95 for the systems displayed, which means that $\langle r^2 \rangle$ is up to 10% smaller for fluoride ions in the solid than in the gas phase. Chloride ions are much larger in extent, due to the larger atomic

number, and are more greatly affected by the solid than the fluoride systems, with up to 20% reductions in $\langle r^2 \rangle$. The stabilization of the divalent oxide anions is even more striking; they do not even exist in the gas-phase. As a comparison, calculations using the crystal-perturbation model on helium at very high compressions (up to six fold over the low-pressure solid) found that $\langle r^2 \rangle / \langle r^2 \rangle_g$ was reduced to about 0.94 at 23 GPa. As in the ionic solids, this compression of the size of the atoms lowered the calculated volume and improved agreement with experiment.

4. CONCLUSIONS

The Gordon-Kim electron-gas model coupled with simple descriptions of the electronic structure of the constituents yields remarkably good descriptions of the structure and energy of simple ionic and molecular solids. By building into the electronic structure of the constituent atoms and ions the dominant changes due to the crystal environment, only the interaction energies need be determined, thus reducing considerably the computational effort. The description of the electronic structure is itself straightforward, making use of the local symmetry in the crystal lattice. For many ionic systems, spherical, unpolarized ions can be assumed with very little error. For others, such as quartz, more complicated models, which include polarization effects, can be used. These methods require only slightly more computational effort. Application to the more complicated molecular systems is straightforward. Experience with the method indicates that the calculated results are very reliable.

ACKNOWLEDGEMENT

The work of Richard LeSar was performed under the auspices of the U. S. Department of Energy and was supported in part by the Division of Materials Science of the Office of Basic Energy Sciences of the Department of Energy. R. L. would like to thank the Departments of Chemistry and Materials Science & Engineering of the University of Michigan for their

support and hospitality. Roy Gordon wishes to thank the National Science Foundation for support.

REFERENCES

[1] P. Hohenberg and W. Kohn, Phys. Rev. 136, 864 (1964).

[2] W. Kohn and L. J. Sham, Phys. Rev. 140, 1133 (1965).

[3] R. G. Gordon and Y. S. Kim, J. Chem. Phys. 56, 3122 (1972).

[4] J.C. Slater, Quantum Theory of Atomic Structure, Vol. II (McGraw-Hill, N.Y., 1960), pp. 323-332

[5] M. J. Clugston, Adv. Phys. 27, 893 (1978).

[6] M. Waldman and R. G. Gordon, J. Chem. Phys. 71, 1325 (1979).

[7] J. Mehl, R. J. Hemley, L.L. Boyer, Phys. Rev. B. 33, 8685 (1986).

[8] L. Hedin, B. I. Lundqvist, J. Phys. Chem. 4, 2064 (1971).

[9] E. W. Pearson and R. G. Gordon, J. Chem. Phys. 82, 881 (1985).

[10] Y. S. Kim and R. G. Gordon, J. Chem. Phys. 61, 1 (1974).

[11] M. Waldman and R. G. Gordon, J. Chem. Phys. 71, 1340 (1979).

[12] J. S. Cohen and R. T. Pack, J. Chem. Phys. 61, 2362 (1974).

[13] R. LeSar and R. G. Gordon, J. Chem. Phys. 78, 4991 (1983).

[14] R. LeSar, J. Phys. Chem. 88, 4272 (1984).

[15] R. LeSar and M. S. Shaw, J. Chem. Phys. 84, 5479 (1986).

[16] C. P. Wood and N. C. Pyper, Mol. Phys. 43, 1371 (1981).

[17] C. Muhlausen and R. G. Gordon, Phys. Rev. B. 23, 900 (1981).

[18] R. LeSar and R. G. Gordon, Phys. Rev. B. 25, 7221 (1982).

[19] Y. S. Kim and R. G. Gordon, Phys. Rev. B. 9, 3548 (1974).

[20] A. J. Cohen and R. G. Gordon, Phys. Rev. B. 12, 3228 (1975).

[21] R. E. Watson, Phys. Rev. 111, 1108 (1958).

[22] A. Cohen and R. G. Gordon, Phys. Rev. B. 14, 4593 (1976).

[23] M. D. Jackson and R. G. Gordon, Phys. Rev. B. 38, 5654, (1988)

[24] G. H. Wolf and M. S. T. Bukowinski, Phys. Chem. Minerals 15, 209 (1988).

[25] E. Pachalis and A. Weiss, Theor. Chim. Acta 13, 381 (1969).

[26] R. J. Hemley and R. G. Gordon, J. Geophysical Res. 90, 7803 (1985).

[27] R. J. Hemley, M. D. Jackson and R. G. Gordon, Geophys. Res. Lett. 12, 247 (1985).

[28] L. L. Boyer, M. J. Klein, J. L. Feldman, J. R. Hardy, J. W. Flocken and C. Y. Fong, Phys. Rev. Lett. 54, 1940 (1985).

[29] M. D. Jackson and Roy G. Gordon, unpublished.

[30] H. Hummel, D. Lacks and R. G. Gordon, unpublished.

[31] M. D. Jackson and R. G. Gordon, Phys. Chem. Minerals 16, 212 (1988).

[32] M. D. Jackson and G. V. Gibbs, J. Phys. Chem. 92, 540 (1988).

[33] J. E. Post and C. W. Burnham, American Minerologist 71, 142 (1986).

[34] M. D. Jackson and R. G. Gordon, Phys. Chem. Minerals 15, 514 (1988).

[35] R. LeSar and R. G. Gordon, J. Chem. Phys. 77, 3682 (1982).

[36] P. W. Fowler and M. L. Klein, J. Chem. Phys. 85, 3913 (1986).

[37] R. LeSar, Phys. Rev. B. 28, 6812 (1983).

[38] R. LeSar, Phys. Rev. Lett. 61, 2121 (1988).

[39] G. Schoknecht, Z. Naturforschung 12, 983 (1957)

LDA APPLICATIONS TO THE PROPERTIES OF PERIODIC SYSTEMS

R. C. Albers
A. M. Boring
J. C. Boettger

University of California
Los Alamos National Laboratory
Los Alamos, NM 87545

1. Introduction

Excitement in physics almost always occurs at the fringe - in areas that are not well understood and are typically beyond the limits of applicability for standard theories - in short, at the frontiers of physics. Condensed matter electronic structure theory is no different in this respect from any other area of physics. It should come as no surprise, therefore, that the Local Density Approximation to the Density Functional Theory (LDA-DFT) is often applied to systems or problems that, strictly speaking, are beyond it's known range of validity. The purpose of this review article is to focus on a few of these areas and to show how well or poorly LDA-DFT theory works and what generalizations or extensions seem to be required to successfully explain the physics in these areas. Because of the vastness of modern condensed matter physics, the authors have been forced to neglect many interesting areas (some of which are treated in other articles in this book) and to restrict the discussion to a few bulk and layer applications in which they have some direct involvement or interest.

ADVANCES IN QUANTUM CHEMISTRY
VOLUME 21

Before discussing applications of LDA-DFT, it is useful to compare that theory with Hartree-Fock (HF) theory, which is also based on a variational principle and requires solving a set of nonlinear differential equations for one-electron orbitals and energies. Indeed, the earliest LDA calculations were performed prior to the advent of DFT theory and attempted to simply approximate HF theory. In spite of these similarities there are fundamental theoretical differences, which limit the range of applicability for each of the theories.

HF theory is based upon a single Slater determinant approximation to the ground state wavefunction, which is varied to minimize the energy obtained from the exact many-body Hamiltonian. Thus the HF theory provides a rigorous upper bound to the exact ground state energy, with the error being due to neglect of electron-electron correlation. The fact that HF theory has a well-defined ground state wavefunction, allows one to rigorously interpret the one-electron orbitals and energies as one-electron excitations, within the context of Koopmans' theorem, i.e., neglect of correlation and relaxation. The primary advantage of HF theory lies in it's theoretical rigor for both bulk and one-electron properties and the fact that the errors introduced are, in principal, well-understood. The primary disadvantages are the necessity of solving nonlocal one-electron equations, the relatively large correlation and relaxation errors in the one-electron properties of solids, and because correlation calculations are only practical for fairly small systems. These limitations have generally restricted HF calculations to atoms and molecules.

In contrast, LDA-DFT (which we will refer to as simply LDA) is based upon a local density approximation to the as yet unknown density functional whose variational minimum with respect to density would be the exact ground state energy. Since the exact density functional is not known, the LDA energy

does not represent a rigorous bound for the exact energy and the error introduced is unknown. The primary advantages of LDA are the intrinsic locality of the one-electron equations, errors in the energy which become small in regions where the density is smoothly varying (such as bonding regions in extended systems), and one-electron eigenvalues which are in quite good agreement with experiment for extended systems. The primary disadvantages of LDA are it's lack of theoretical rigor for one-electron properties, relatively large errors in the one-electron eigenvalues for highly localized states, and because the errors introduced into both the bulk and one-electron properties are often poorly understood. As a result of the advantages and limitations just discussed, LDA has become the primary tool applied to extended systems, while HF theory has remained the primary tool for studying atoms and molecules.

From the above discussion of LDA, it should be apparent that the range of application for LDA is poorly defined. Nonetheless, one of the most remarkable results of LDA research over the past 25 years has been the discovery that for a wide range of systems, both the bulk and one-electron properties are often well described by LDA. Indeed, since success has almost become the norm, a great deal of recent work has been devoted toward finding ways to fix up LDA in the few places it fails. This work has been facilitated by two important developments. The first is the development of high-precision full-potential methods for solving the LDA one electron equations (as opposed to using some type of shape-constrained potential and density). The second is the modern treatment of relativistic effects associated with heavy (high-Z) atomic components in solids. Here, instead of solving the four-component Dirac equation for a crystal with many atoms per unit cell, one can decouple the Dirac equation to obtain the mass-velocity and Darwin corrections in a non-relativistic L-S basis. This procedure,

called either the semi- or scalar relativistic technique, does not include spin-orbit coupling, but does allow spin-polarization to be included for high Z components and much of the published literature is based upon this method. However, with modern supercomputers and new numerical algorithms results using the Dirac equation have also become fairly widespread. Recently, methods for rigorously including both spin-polarization and spin-orbit coupling have been developed, but are not standard in most LDA applications/1/.

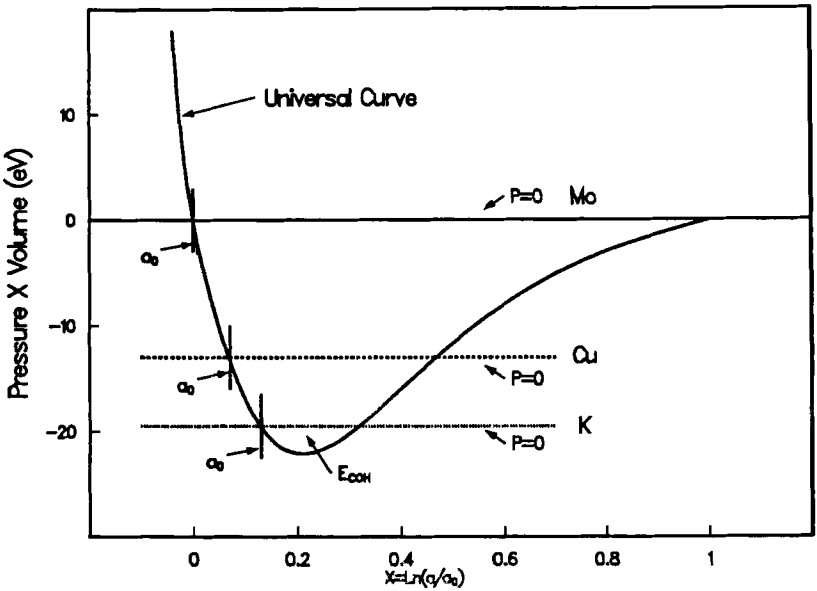
In a sense, looking for the boundary where LDA breaks down is a theme of this review. Specifically, we discuss first the bulk calculations of cohesive energies, bulk moduli, and lattice constants. Next a very brief discussion of elastic constant calculations is given. The last of the properties based on the total energies that we discuss is that of crystal phase stability. We then look at comparisons between the LDA one-particle spectra and various measured quasi-particle spectra. Comparisons are made with x-ray photoemission (XPS) spectra and optical measurements. Next we focus on the determination of those states in the small energy range around the Fermi Surface (FS). Finally we examine complex systems such as the Heavy Electron systems or low dimensional structures, where one is again pushing the limits of the LDA-DFT method.

2. Cohesive Energies, Lattice Constants and Bulk Moduli

The calculation of these properties are the least demanding of LDA calculations in the sense that one can calculate the cohesion, lattice constant, and bulk modulus of all the d and f band elements of the periodic table to within 5-10% of the experimental values treating them all as fcc crystals/2/. That this is true of the d band metals and the rare-earth metals

(where, with the exception of Ce, the f electrons are in localized states and so they are d-band metals) is not surprising since the ground states are usually fcc, bcc, hcp, or dhcp structures for which the energy differences are small. That this is true of the light actinide metals (Pa - Pu) with low symmetry ground states is surprising/3/. The parabolic trend of cohesion of the transition metals can be understood in terms of having a maximum number of unpaired bonding d orbitals filled in the middle of the series. However, this simple model does not explain many of the more subtle details. For example, in this simple picture K and Cu should have roughly the same cohesion due to 4s orbital bonding. The LDA calculations show that the cohesion in Cu is five times greater not only because of a d orbital contribution (20%) but because the s-d hybridization makes the 4s bonding stronger in Cu. These calculations also show that the rare-earth and possibly the actinides metals are d-bonding.

This information is captured in the universal bonding curve of the transition metals, where the force (or pressure) on a unit cell of the crystal is calculated as a function of interparticle distance/4/. The generated curves are based upon LDA calculations in which the atoms are pulled apart while maintaining the crystal symmetry. In Fig. 1, a version of the universal bonding curve illustrating the behavior of a transition (Mo), a noble (Cu), and an alkali (K) metal is given. This example shows, that if a given metal has a larger cohesion than another, then its bulk modulus (slope of curve at a_o) is larger and its lattice constant is smaller based upon its position on the universal curve. From these LDA calculations we know that most of these metals have a broad sp-band that hybridizes with a narrower d-band and that near equilibrium the d-band pressure is always



attractive and the sp-band pressure is always negative. Thus the balance between these two pressures yields the observed lattice constants. We also know that repulsive sp-band pressure is the dominant contribution to the bulk moduli. This gives us a picture of metallic bonding as a hybridized sp-d bond with weak directional character yielding close packed structures for all the alkaline, noble, transition, rare-earths, and actinide metals (some 65 elements). The only exceptions are the light actinides where the d-f hybridization generates non-close-packed structures.

For compounds the situation is more complicated and here we give a very brief review of this field. Again, the computed results for lattice constants, cohesion, and bulk moduli are within 10% of the experimental values for real materials. Because of charge transfer effects with different atoms on different sites, the analysis of bonding is not so straight forward as it is for metals. (The assignment of charge in a

multi-site problem with overlapping orbitals is ambiguous.) As an example, we show in Table 1 the calculated results for Al, Ni, and Ni_3Al all calculated in the same manner. As can be seen, the calculated results for the compound are as good as for the elements from which it is formed/8/.

TABLE 1. LMTO RESULTS

AL(FCC)	LMTO	EXPT.	DEV.
a_0(Å)	4.0345	4.0498	0.4%
B(Mbar)	0.82	0.79	3.0%
E_{coh}(eV)	3.74	3.36	9.0%
Ni(fcc)	LMTO	EXPT.	DEV.
a_0(Å)	3.5141	3.5198	0.2%
B(Mbar)	1.99	1.80	9.5%
E_{coh}(eV)	4.45	4.45	0.0%
$Ni_3Al(Cu_3Au)$	LMTO	EXPT.	DEV.
a_0(Å)	3.571	3.5667	0.1%
B(Mbar)	1.78	1.64	7.9%
E_{coh}(eV)	4.79	4.57	4.6%

Another goal of LDA research in the last six years has been to calculate elastic constants by calculating the changes in the total electronic energy associated with a given deformation. Most of this effort has focused on cubic crystals where there are only three independent elastic constants (C_{11}, C_{12}, and C_{44})/6/. To obtain C_{11} and C_{12} one performs a volume-conserving tetragonal distortion, which yields the shear constant $C_s = \frac{1}{2}(C_{11} - C_{12})$. To obtain C_{11} and C_{12} separately one performs a uniform dilation δ. Finally one obtains C_{44} from a volume-conserving trigonal distortion along the (111) axis of the cube. The elastic energy is then equated to the change in the total electronic energy upon distortion.

The calculation of elastic constants has been a strong incentive for the development of full potential methods. For

example, in the earlier linear muffin-tin orbitual (either ASA or its equivalent) calculations the trend in elastic constants for the elements were given correctly, but qualitatively the results were poor/9/. From these early results it was determined that problem was that the contributions from the non-spherical charge density generated upon distortion were too large to be neglected. Recently there have been full potential LDA calculations on both elements and compounds that are in agreement with the experimental elastic constants at the 10 % level. For example, Wills, et. al., has recently calculated the elastic constants for the elements Cu, Ir, and W that are within 10% of the experimental values (except for C_{12} of Cu in which the error was 17%)/10/. An example of compounds calculations is the determination of the elastic constants of NbC and MoN/11/. So, with the full potential techniques one can now obtain the same level of agreement between theory and experiment for elastic constants as was obtained earlier for lattice constants cohesion, and bulk moduli.

3. Crystallographic Phase Stability

In the last ten years, LDA has been successfully applied to the crystallographic phase stability of solids. For example, Skriver/12/ recently studied the low pressure phase stability of 42 monatomic metals and correctly predicted the stable structure for 35 of those metals. LDA has also been successfully calculated the sequence of pressure-induced phase transitions for a number of solids/13-17/, including several instances in which the theoretical calculations correctly predicted transitions that only later were detected experimentally/14,15/.

These successes are impressive and not something that one would have anticipated, since the structural energy differences

are small compared to the total energies. For example, one LDA estimate of the intal atomic energy for Mg is more than a Ry away from the more accurate Hartree-Fock value/5/. In contrast, the energy differences between the hcp, fcc, and bcc phases of Mg are a few mRy/10/, a difference of 3 orders of magnitude smaller than the total energy. Nonetheless, LDA calculations by McMahan and Moriarty on hcp, fcc, and bcc Mg not only reproduced the known crystal structure at zero pressure (hcp), but also predicted a hcp to bcc phase transition between 50-57 GPa, which was subsequently detected at 50 GPa/11/.

To understand this success, note that LDA is nearly exact for a homogeneous electron gas and has errors that increase as a system becomes more inhomogeneous. Also, the errors in LDA electronic band structures for metals tend to be small near the Fermi level and progressively get larger for lower lying occupied states. Hence LDA errors should be larger for core states than for bonding states. This is in sharp contrast to HF theory, which produces substantially better total energies, but concentrates it's error (neglect of correlation) in the bonding regions.

The importance of valence bonding or band-structure effects in determining crystal structures manifests itself in a number of ways. One of the most general of these is an apparent relationship between pressure induced electronic and structural transitions in solids. McMahan/15/ has reviewed a number of standard sequences of phase transitions which are related to an s to d electronic transition under pressure. More recently, it was shown that a 2s to 2p electronic transition in Li should induce a series of structural transitions (fcc -> hcp -> bcc) for pressures above 5 Mbar/17/. A number of the other manifestations of band structure effects on crystalline phase stability have been reviewed by Skriver/12/.

The development of high precision techniques for solving the LDA equations has transformed phase stability research into an important tool for probing the boundaries of LDA. Several recent calculations have revealed examples of systems where LDA clearly predicts the wrong crystal structure at zero pressure/12,18-20/. The best studied example of this type of discrepancy is that for Fe/18-20/. Recent calculations show unambiguously that LDA predicts fcc instead of bcc stability for Fe, in clear disagreement with experiment/18-20/. The present state of the art for LDA band structure calculations is such that Fe should provide a good test for future extensions of DFT beyond the LDA approximation.

4. X-RAY PHOTOEMISSION SPECTRA (ANGLE-INTEGRATED)

In photoemission experiments one shines light (incident photons) on the sample that excites the electrons (via the electron-photon interaction) and then analyses the energy of the ejected electrons. For a review of the details of these experiments see Ref. 21. Here we will give a brief outline of the LDA theoretical model used to understand results of these experiments. The theory can be based on the work of Caroli et al.,/22/ yielding an exact many-body photoemission formulation which includes electron-electron interactions to any order in perturbation theory. Taking this theory to lowest order, which corresponds to an independent-electron approximation, Feibelman and Eastman/23/ have converted it into a Fermi's Golden-Rule-like formula in which the final-state electron has a wavefunction corresponding to that used to describe a low-energy-electron diffraction (LEED) experiment. (Here the collected electron has a velocity which is minus that of the LEED Beam.) To develop a tractable theory one starts with the expression for the radial photocurrent J_r,/23/

$$R^2 J_r = K(E + \omega) \iint dr dr' \phi^+ H_1^+ \mathrm{Im}G\, H_1 \phi$$

with $K(E + \omega) = -\frac{8}{\pi}\left[\frac{1}{4\pi e(E+\omega)}\right]^2 \sqrt{2(E+\omega)}$. In this expression H_1 is the electron-photon interaction, $G(r,r',E)$ is the one-particle Green's function of the initial state (at energy E), $\phi\,(r,R,E+\omega)$ is the time-reversed LEED function, and $\hbar\omega$ is the incident photon energy, and $\hat{R}$ is the distance from the crystal to the detector. Physically, ϕ is the wavefunction of the ejected electron and G contains all of the multiple-scattering effects of the initial-state wavefunction (DOS function). At this point ϕ includes all the multiple-scattering effects of the ejected electron in transversing the crystal. (In XPS experiments the incident photon energy is around 1.5 keV and so the escape depth for the electrons can be about 50 Å.) Winter et al./24/ have shown that at KeV energies the rapidly oscillating phase factors suppress coherent multiple-scattering and a single-site approximation for ϕ can be made. Next, using the periodicity of the solid and restricting H_1 to be the electric dipole operator, we obtain an interaction proportional to the gradient of the crystal potential. Remembering that the partial local DOS functions (n_k) can be expressed in terms of the imaginary part of the Green's function, we now have J_r as a product of a single-site ϕ, the gradient of potential at that site, and the partial local DOS function. The angle-integrated photocurrent is given by an integration over $\hat{R}$,

$$I(E + \omega) = \int R^2 J_r(R, E + \omega) d\hat{R}$$

and with a little bit of algebra one obtains the final expression (given here for the relativistic case)/20/

$$I(E + \omega) = \sum_{\text{sites}} \sum_{\kappa} \sigma_{\kappa}^{i}(E + \omega)\, n_{\kappa}^{i}(E)$$

This final expression has been cast into a Fermi-golden-rule form, where $n_{\kappa}^{i}(E)$ are the partial DOS functions for the κth channel on the ith site, and are given by a standard LDA calculation. The relativistic cross-section $\sigma_{\kappa}^{i}(E + \omega)$ are calculated over the energy E of the DOS functions for a fixed incident photon energy $\hbar\omega$ (usually around 1.5 keV).

The importance of performing these calculations is that one can then make a direct comparison with experimental data. To illustrate this model we discuss the published results on UPt_3 and UBe_{13}/25,26/. First of all, the calculated and measured XPS intensities in UPt_3 are in very close agreement. Next this example shows that the energy dependence of the cross-sections of a given κ (or ℓ) is inversely proportional to bandwidth of that κ value. (So very wide bands have almost a constant cross-section.) In the UBe_{13} example one sees the importance of the relative values of cross-sections of the different atoms of the system. Here the Be DOS function (13 atoms) is as large at E_F as the f DOS function from the single U atom, but the Be (ℓ=1) cross-sections are 30 times smaller and experimentally there is no Be component at E_F. The calculated XPS intensities also show no Be component at E_F. The agreement between these one-electron calculations and XPS data is not perfect, indicating many-body effects. As more and more of these calculations are performed they should bring into sharper focus the many-body effects in real systems.

5. Optical Measurements

The interaction of light with matter is usually weak (in the absence of high powered lasers). Hence, the optical absorption for interband transitions, which is proportional to the imaginary part of the dielectric function $\epsilon_2(\omega)$, can also be calculated from a simple golden rule expression/27/. In the frequency range of interest for optical transitions (on the order of eV), the photon wavevector is much smaller than the Brillouin zone size and hence is ignored (dipole approximation) and only direct transitions are considered. In the early days, when computer power was limited, the joint density of states (i.e., setting the matrix element squared to unity) was often compared to optical spectra. While this gives some feeling for the optical spectra, the energy dependence of the matrix elements is needed for a better comparison with experiment. In metals one must add a low frequency intraband absorption (Drude correction) to the optical spectra/27/ and in semiconductors and ionic compounds one must also consider excitons/22/ and possibly impurity absorption as well.

The implicit assumption in all of this is that the LDA eigenvalues are the same as quasiparticle eigenvalues. While this should not be true in general, in practice the good agreement that is often achieved suggests that this is a reasonable approximation for metals. Early on, good agreement was found for Al and other free-electron-like metals using perturbative pseudopotential theory/28/. When more general band-structure methods such as the linear combination of gaussian orbitals (LCGO) or linear muffin-tin orbitals (LMTO) approaches are used (see references cited in Ref. 29) on more complicated metals, e.g., FeAl or ferromagnetic Fe, good agreement is still achieved. Older calculations that did not trouble to tackle the complications of accurate double Brillouin

zone integrations by modern methods such as the tetrahedron method, which a good optical calculation demands, nonetheless found good agreement with optical spectra for many of the transition metals by examining the bands themselves and looking for likely singularities in the joint density of states that should cause corresponding peaks in the experimental optical spectra. A detailed comparison by such means may be found in the review article by Macintosh and Andersen/30/.

In semiconductors the chief problem is that LDA typically underestimates gaps by as much as 50%, or even predicts metals or semimetals instead of insulating behavior (i.e., closes the indirect gap). At least two methods have been used to overcome this problem. The most often used is an empirical pseudopotential that is adjusted until the experimental gap is correctly calculated/27/. Another approach that has been recently proposed/31/ is to add an additional repulsive broadened delta-function-like external potential around the nucleus of each atom. The width and magnitude of this potential are adjusted until the gaps at the Γ, X, and L points (for rocksalt structures) fit experiment. In effect, an additional artificial term has been added that mimics the Darwin term in relativistic calculation. Because of its short range, its main effect is to shift only the s-like bands up in energy and thereby opening up the gaps. Once the gaps at one lattice constant are correctly adjusted, the volume-pressure dependence of the optical spectra seems to be correctly calculated without further adjustment of the external potential.

The optical measurements probe occupied and unoccupied electronic structure on the eV scale of energy. Another technique, x-ray absorption spectroscopy (XAS) probes the unoccupied states up to about a keV above the Fermi energy. In this spectroscopy x-rays are absorbed by electronic transitions from atomic core levels to the unoccupied states above the Fermi

energy. Dipole selection rules lead to projections of the electronic structure onto specific angular momentum components. Special band-structure techniques have been developed by Müller and Wilkins to calculate this type of spectra/32/. These techniques reflect the difficulties in accurately calculating electronic states with high kinetic energies; they require much larger basis sets with many more plane waves and angular momentum basis functions. In addition, linearized methods run into difficulties because of the many overlapping bands of different angular momentum and the large energy range over which the calculations must be done. Although the XAS data extends up to a keV above threshold, in practice the technical difficulties mentioned above will probably restrict band-structure calculations to a few hundred eV above the Fermi energy. The higher energy spectra is also considerably reduced in amplitude by Debye-Waller type effects and hence is probably not worth pursuing very strongly in any case. Although there is considerable core-hole and final-state broadening that washes out XAS, there is good agreement between these types of calculations and experiment in predicting the relative positions and intensities of every peak and valley in the metal K-absorption edges out to 200 eV above threshold/33/. The only caveat is that the experimental energy scale must be slightly compressed/34/. It is unclear why this is necessary, but Rehr/35/ has done calculations that suggest that this may be corrected by an energy-dependent exchange-correlation potential.

6. Fermi Surfaces (Topologies and Mass Enhancements)

Fermi surface calculations have always been an important test for LDA band-structure calculations. The existence of good de Haas-van Aphen (dHvA) experimental data for the extremal

orbit areas provides crucial experimental verification of the validity of the one-electron picture and of the various approximations used for the potentials. In this section we will only briefly review this area for nonmagnetic metals. For more details, an excellent discussion of dHvA experiments and Fermi surface properties is given in Ref. 36 and the situation for the more complicated itinerant magnetic systems in Ref. 37.

The most fundamental problem facing LDA Fermi surface calculations is that the theory in this case has no sound theoretical underpinning. The standard LDA procedure is to solve the equation $\epsilon_{n,\vec{k}} = \mu$, where the $\epsilon_{n,\vec{k}}$ (for wave vector $\vec{k}$ and band index n) are the one-electron eigenvalue spectra obtained by solving the Kohn-Sham equations and μ is the chemical potential. Since the eigenvalues in LDA have no obvious meaning, this approach is suspect. The correct prescription is to solve the quasiparticle Dyson equation for the quasiparticle energies $E_{n,\vec{k}}$ and then search for the $\vec{k}$ that satisfy $E_{n,\vec{k}} = \mu$. In practice, since many LDA calculations have been performed that are in reasonable agreement with experimental FS's, it is clear that the LDA eigenvalues must in some way mimic the true quasiparticle energies. This can be justified by viewing the local exchange-correlation potential $v_{xc}(\vec{r})$ as an approximation for the mass operator, i.e.,

$$\int d\vec{r}' M(\vec{r},\vec{r}'; E_{n,\vec{k}})\psi_{n,\vec{k}}(\vec{r}') \simeq v_{xc}(\vec{r})\psi_{n,\vec{k}}(\vec{r}) \quad .$$

Although this is attractive, it is hardly rigorous and probably not strictly correct.

Since the 1960's when electronic band structure methods became sophisticated enough and available computers fast enough for the first quantitive FS calculations on non-free-electron-like materials, three controversies have

continually arisen: (1) the quality of the LDA potential used, (2) the quality of the band-structure methods themselves, and (3) whether a non-local exchange-correlation potential is required. The first two controversies, over the potentials and methods, arose because calculations were initially non-self-consistent and used a muffin-tin approximation for the potentials; a Mattheis procedure that overlapped atomic charge densities was used to construct the potential. As the first self-consistent potentials started appearing, further controversy was generated over which exchange-correlation potential to use. Finally, when full potential methods began to be used, there was great interest in non-muffin-tin corrections to the FS. At the present time, with self-consistent full potential methods we have gone about as far as possible in solving the first two points without further fundamental advances in the formalism. In retrospect, these new advances, while important for sorting out where the problems really were, often do not provide a FS that can fit experiment any better than the original non-self-consistent calculations. The end conclusion, point (3), seems to be that a non-local exchange-correlation potential is required to get the correct Fermi surface and even this may not be enough: it is still an open question whether or not one is forced to solve explicitly for quasiparticle energies in order to remove the small remaining discrepancies between the calculated and experimental Fermi surfaces. There exists at least one counterexample model calculation that shows that the LDA eigenvalues are not equal to the quasiparticle energies/38/.

The question of nonlocality has been most closely examined for free-electron-like materials. The overall Fermi surface for these types of materials is almost forced to be reasonable, since the Fermi surface must be roughly spherical and Luttinger's theorem determines the volume enclosed by the Fermi

surface. However, LDA calculates these small deviations from spherical behavior to be twice as big as are observed experimentally. Non-local corrections seem to help reduce these deviations/39/.

For noble, alkaline-earth, and transition metals/30/, and even for strongly correlated materials/40/, the LDA Fermi surface is often in good agreement with measured dHvA external orbit areas, often predicting areas within 10% of experiment. In order to achieve such good results, however, it has often been found necessary to shift either the Fermi energy or the bands by a small amount. These shifts are usually rationalized by an appeal to the non-local exchange-correlation parts of the potential that are left out of the LDA. Based on experience with nonlocal pseudopotentials used in semiconductors/15/, it is known that non-local effects can be mimicked by an ℓ-(or orbital angular momentum-) dependent potential. If such a potential were to be used in a band-structure calculation, the net effect would be to shift bands by amounts depending on the differences between the ℓ-dependent potentials weighted by the projected ℓ-dependent character of the bands. For transition and noble metals shifts on the order of 10-20 mRy of the sp bands relative to the d bands are typically needed to bring theory into agreement with experiment/41/.

Another aspect measured by dHvA experiments is the effective mass of the electrons, averaged over the various extremal orbits. In contrast to the Fermi surface itself, which is not expected to show many-body corrections/42/, the electron effective masses should be enhanced over the band structure effective mass by a factor of $1 + \lambda$ for the electron-phonon interaction/43/ and a similar factor for the electron-electron interaction. In Table 11.1, p. 256, of Ref. 43, Grimvall gives a list of recommended electron-phonon mass enhancement parameters, λ, for a large number of metals. These range from

0.13 for K to 1.55 for Pb. In Au the average enhancement/26/ over the different extremal orbits as measured by comparing band calculations to dHvA is about 0.10. In many materials the electron-electron enhancement is considered to be very small and is often not estimated. The one striking exception is heavy-fermion systems, where these enhancements dominate the electron-phonon enhancements. In UPt_3, e.g., the enhancements over the different extremal orbits/40/ (i.e., the experimental mass divided by the band mass) range from 10-30, with 20 being a typical enhancement. These enhancements agree well with the specific heat enhancements/44/ of about 22.

7. LDA Calculations of Many-Body Parameters

A relatively new use for LDA calculations that is becoming increasingly important in understanding systems with strong electron-electron correlations is to estimate parameters for model many-body Hamiltonians/45/, which are sufficiently simple that sophisticated many-body techniques are available to provide a better treatment of the correlations than the mean-field LDA approach can do. An example of such a model Hamiltonian is the two-band Hubbard model that is being applied to the two-dimensional Cu-O planes of the new high-temperature superconducting materials (e.g., La_2CuO_4), which can be schematically written (we ignore all site and spin indices):

$$H = \sum \epsilon_p n_p + \sum \epsilon_d n_d + t \sum_{nn} (p^\dagger d + p d^\dagger) + U_d \sum n_d n_d + U_p \sum n_p n_p + V_{dp} \sum n_d n_p \ .$$

The parameters ϵ_p and ϵ_p are the band centers of the tight-binding O p and Cu d bands respectively. In the schematic

Hamiltonian we only allow hopping between nearest-neighbor O p and Cu d sites, which is controlled by the hopping parameter t. More generally, other hopping terms can also be added. The Hubbard U parameters U_d and U_p control the size of the onsite Coulomb repulsion terms. The V_{pd} parameter controls the nearest-neighbor Coulomb repulsion between the d and p orbitals. Longer-range Coulomb terms could also be added, but it is fairly standard to stop at the nearest-neighbor level. The other standard model Hamiltonian of interest is the impurity Anderson Hamiltonian, which can be schematically written as (again ignoring site and spin indices):

$$H = \sum_k \epsilon_k \, c_k^\dagger c_k + \sum V_{fk}(c_k^\dagger f + f^\dagger c_k) + U_f n_f n_f \quad .$$

This Hamiltonian, which was originally developed to understand local moment formation, may be viewed as a localized orbital (here represented by an f orbital) that hybridizes with matrix elements V_{fk} with a set of band states with band energies ϵ_k. Again the Coulomb term involving U_f is responsible for the strong on-site correlations. Without this term the Hamiltonian is biquadratic and can be exactly diagonalized. The Anderson-type Hamiltonian has most recently been applied to mixed-valence, Kondo-impurity, and (when generalized to contain a lattice of impurity sites) heavy-fermion systems. Both of the above model Hamiltonians are, of course, extreme approximations to the real systems; there are many more terms that should be included. The conventional hope is that whatever interactions are neglected will not change the physics of the standard simple models, but will simply renormalize the standard parameters (i.e., their effects can be adequately represented by changing the parameter values).

Because these model Hamiltonians represent an Ansatz and are not derived from first principles, except in an intuitive sense, there is some ambiguity in how to calculate their parameters. If we think of the different orbitals as real localized orbitals (actually, Wannier states, since most of the model Hamiltonians implicitly assume orthogonal states), then the various hopping parameters and Coulomb repulsion terms could be calculated directly in terms of the wave functions of the orbitals. Different size orbitals would then give different sets of parameters. In practice, because of the implicit renormalizations and orthogonalizations and the drastically simplified nature of the Hamiltonians, LDA calculations of the parameters usually assume that the quasi-localized orbitals must somewhat resemble the Wannier functions that make up the band states. Thus the hopping parameters, hybridization matrix elements, and band centers are typically derived from tight-binding fits to the LDA band states. To make the calculations tractable it is further assumed that these parameters are independent of the Coulomb correlations. The most difficult part of the calculation is the Hubbard U (or onsite Coulomb) parameters. These parameters are calculated by constrained total energy calculations. The orbital of interest, which we will call a d orbital, is put in as a core orbital that does not hybridize with any other states. In the simplest treatments all of the other electrons are then allowed to self-consistently relax as a function of the number (n) of d electrons; for this reason the calculated U's are said to be screened parameters. The total energy is then fit to the form: $E(n) = \text{constant} + n\epsilon_0 + (1/2)Un(n-1)$. Alternatively, one can use a theorem of density functional theory that the one-electron eigenvalue ϵ_d is given by $\partial E(n)/\partial n$, or in our case by $\epsilon_0 + U(n - 1/2)$. To obtain U_{pd} type parameters, one must similarly vary two different occupation numbers. To handle charging effects

due to adding or subtracting electrons one can either force the unit cell to be neutral (in effect, transferring electrons between the localized orbitals and the other band states) or by adding a uniform neutralizing positive charge background. Ideally, the localized orbitals should be thought of as being on impurity atoms that are embedded in the otherwise perfect lattice. In practice, supercells are often used that have a lattice of such impurity atoms with the hope that electron screening will reduce any significant impurity-impurity interactions. This is often an excellent approximation and for some systems the screening seems to localize the effects of the occupation changes to within the original unit cell so that even supercells are not necessary. Another condition to make these procedures meaningful is that the localized orbital should actually be localized. If the orbital leaks out of the atomic sphere radius by any significant extent, a core-wave-function representation of the localized orbital becomes less meaningful. This problem presents itself in the copper-oxide high-temperature superconductors, since the oxygen p-orbitals are quite extended.

8. Heavy Electron Systems

Heavy-electron (sometimes called heavy-fermion) systems are a particularly striking example of the partial success that LDA theory can achieve even when applied to some very unlikely systems. The heavy-electron systems are cerium or actinide compounds (i.e., materials containing atoms with a partially filled f shell) that have enormous specific heat and magnetic susceptibility enhancements that are believed to arise from very strong electron-electron correlations. Useful reviews of the experimental literature for these systems are given in Refs. 41, and for the various theoretical approaches in Refs. 42 and 43.

An LDA band structure approach to these systems has often been characterized as irrelevant. This dismissal has been based on both conceptual and practical grounds. Typical conceptual grounds include, e.g., (1) strong electron-electron correlations are not expected to be handled very well in LDA approaches; (2) large entropies were found experimentally for some of these systems that would presumably imply localized excitations instead of an itinerant band structure; (3) the success of the Kondo model in describing magnetic impurities in metals suggested to many that a "Kondo lattice" is a better description of heavy-electron behavior; and (4) strong magnetic fluctuations that are beyond LDA theory have been observed in some systems. The most obvious practical grounds for dismissal are the huge experimental linear specific heat coefficients, which are often 20-30 times larger than the LDA predictions.

Many of these objections are probably correct and there is no doubt that a complete theory of heavy-electron behavior requires a theory with much more physics in it than can be encompassed within a conventional LDA. This is neither disturbing nor surprising. Much of the anomalous properties of heavy-electron behavior are a consequence of the unusual low energy excitation spectrum due to the strong electron-electron correlations, whereas LDA theory has always been expected to only be strictly correct for ground state properties.

When we consider experiments that measure ground-state-like properties or ignore the low energy excitation spectra, LDA methods do quite well. As mentioned earlier, XPS experiments that measure the occupied electronic states are in quite good agreement with LDA results for UPt_3. Even more strikingly, LDA band calculations/40/ for the Fermi surface of UPt_3 are in excellent agreement with dHvA/49/.

In terms of experiments that probe the electronic structure away from the Fermi surface, valence-band XPS experiments are

the simplest to interpret because they are a high-energy bulk measurement. For the heavy electron systems UPt_3 and UBe_{13} good agreement is achieved between theory and experiment/25,26/. However, it should be noted that the agreement is much better for UPt_3 where the spectra mainly shows the filled Pt d bands than for UBe_{13} where the U f bands dominate. Allen et. al./50/ have done resonant PES and BIS spectra, which they interpret as showing U f spectral weight over a much larger energy range than the LDA U f bands. This claim is difficult to validate because PES spectra strongly probe the surface region, whose lower symmetry geometries would tend to widen the f bands, and also involve problems with final state effects (by this we mean that electronic structure in the 100 eV region above the Fermi energy must also be folded in at least at the level of a joint density of states for a simple interpretation of the PES spectra). On the other hand, the PES off resonance looks very similar to the XPS data, which suggests that surface and other effects are not important for this system. This leads to the following problem: band structure seems to give good agreement with XPS data but it produces f bands that are far too narrow to explain the PES and BIS data. This problem has been noticed by Allen in a recent paper/51/, who suggests that the poor resolution of the XPS experiments is the cause of the problem. We believe that the XPS data probably agrees so well because it mainly probes the Pt d bands, which should be as well behaved as any normal transition-metal compound. The U f states, which are almost certainly itinerant, based on the Fermi surface work, do provide important structure near the Fermi energy. However, since they blend into the top of Pt d bands and because of the poor (on the order of 1 eV) XPS resolution it is difficult to separate out their true width. The resonant PES has the advantage of specifically picking out f character.

The huge specific heats and effective masses measured in dHvA experiments have been problematic for an LDA theory from the very beginning. If one accounts for the specific heat by using an enormous static DOS value at the Fermi surface, then the average bandwidth would have to be extremely narrow. Theoretically, such a huge DOS would immediately lead to instabilities; the system would distort to lower the band energy. It seems clear therefore that the effective masses must be dynamic in origin, most likely due to strong low-energy spin fluctuations. This would seem to lie outside the realm of an LDA approach. Nonetheless, in analogy with the quasiparticle mass-enhancement theories of the electron-phonon interaction that is observed in normal metals/42/, some work has been done on "renormalized" band structures. These ideas are reinforced by a number of hints that heavy-electron systems can be described by Fermi liquid theory at low temperatures. Since the LDA already provides reasonable Fermi surfaces, the main idea is to use an ad hoc or phenomenological ("renormalized") f-electron phase shift in the spirit of Landau to increase the effective masses for the f-electron-like quasiparticles and to use the normal LDA phase shifts for all of the other electrons. Zwicknagl/52/, who has reviewed earlier renormalized band approaches, finds that these approaches seem to be able to provide a realistic description of the low-energy excitations in heavy-electron systems without reducing the Fermi surface agreement provided that the spin-orbit splitting is sufficiently large and that crystalline electric field effects are small compared to the quasiparticle bandwidth. From these considerations Ce compounds are expected to show more discrepancies between the quasiparticle and LDA Fermi surfaces than uranium heavy-electron systems.

9. Films and Surfaces

The study of systems exhibiting two-dimensional periodicity provides a perfect example of an area of research for which theory, experiment, and technology, have undergone a rapid and mutually supporting expansion. Under the constant driving force of the technological need to develop ever smaller and faster electronic devices, techniques have been developed which make it possible to build devices or laboratory samples by depositing single layers of atoms on a substrate (or previous layer) in a well-controlled manner. Simultaneously, a number of relatively new experimental techniques have been developed which allow such samples to be characterized and studied in great detail (for example, see Ref. 53). LDA has played a significant role in this expansion by providing the only theoretical tool capable of studying the full range of properties that may be exhibited by surfaces, films, interfaces, and overlayers.

Two-dimensional systems are also a severe test for DFT theory, since such systems combine all of the difficulties inherent in bulk calculations with those of large cluster calculations. Theoretically, there are reasons to expect that the LDA approximation will eventually prove inadequate to deal with many important aspects of these systems such as the inhomogeneity of the Friedel oscillations known to form at surfaces/54,55/, the highly localized nature of some surface states/56,57/, and effects due to image potentials/57,58/. Thus these systems should provide an important test for extending DFT beyond the LDA-DFT approximation.

It should be apparent that the general field of 2d systems is extremely large and we will therefore, only have space to present a few examples of current work that illustrate the diversity of the work being performed and some of the

connections between theory and experiment. For a broader view of some of the older work in this area, the reader is referred to the 1982 review article by Inglesfield/59/.

The work function is probably the most easily experimentally determined property of a film or surface. It is also the only one-electron eigenvalue produced in a DFT calculation which is rigorously interpretable as an electron removal energy/60/. Thus, the work function is one quantity for which extensive comparison between theory and experiment exists. At the simplest level, systematic comparisons have been made between the theoretical work functions for alkali metal and alkaline-earth metal monolayers and corresponding bulk experimental values/61,62/. The theoretical calculations do an excellent job of revealing the qualitative trends exhibited by the experimental data. Quantitatively, the theoretical values are all larger than the experimental values by a few tenths of an eV. This result is not surprising given that the monolayer can be expected to have states which extend further into the vacuum than those for a true semi-infinite solid, and the dipole barrier will be larger.

More extensive LDA calculations have studied the behavior of the work function as a function of film thickness/63-66/. These studies have revealed significant oscillations in the work function as layers are added. For example, Feibelman and Hamann/65/ used LDA electronic structure calculations to predict the work functions of Rh, Cr, Au, and Al for films ranging from 1 to 7 layers thick. In all cases, the work function of the monolayer is a few tenths of an eV larger than the experimentally observed bulk value. As layers are added, the work functions exhibit damped oscillations which, for three of the four metals, converge on the experimental values (see Table II). These predicted oscillations in the work functions of metals represent one example of what are refered to as quantum

size effects (QSE) in thin films. The QSE is a natural consequence of the additional length scale (thickness) that appears in a variety of experimentally observable properties of thin films/67/ when the thickness of the film is comparable to the deBroglie wavelength of the electrons. Thus far the quantitive predictions for QSE in the work functions of thin films have not been tested experimentally due to difficulties in trying to replicate, in the laboratory, the free standing films considered by the calculations/65,67/. Such a test would be quite illuminating, since the QSE in the work function should be highly sensitive to the Friedel oscillations in the charge density.

TABLE II. Work function ϕ (in eV), vs. layer number for various l-layer films and for experiment (taken from Ref. 65)

Film	l = 1	2	3	4	5	6	7	Expt.
Rh(111)	6.51	5.40	5.79	5.59	5.64		5.63	5.1
Cr(001)	4.93	4.46	4.60	4.51	4.48			4.46
Au(100)	5.96	6.13	5.92	5.41	5.66	5.43		5.47
Al(111)	4.74	4.53	4.10	4.43	4.34	4.31		4.26

Another area of interest has been the description of the surface states (SS) which are known to form in the gaps of the bulk electronic band structure/68/. Indeed, a number of LDA calculations for thin films have produced SS which are in good qualitative agreement with experiment/56,59,69-73/, i.e., they produce peaks in the photoemission spectra at roughly the correct energies. However, it is not yet clear that the SS produced in LDA calculations are capable of giving detailed quantitative agreement with more recent angle-resolved photo emission spectroscopy (ARPES) data. For example, the

(001) surface of W exhibits several SS in both ARPES data (see Ref. 59) and LDA calculations/74/. However, the actual shapes of the rather complicated SS bands are noticeably different. For Cu (110), the situation is even worse with ARPES data revealing the existence of SS which do not appear in the calculated bands and failing to detect those which do appear in the calculations/75/. It is not clear if these difficulties are due to fundamental problems in LDA or if they are a result of other factors, i.e., surface reconstructions, finite thickness effects in the calculations, etc.

One example of a SS band which has a simple shape, and which has been mapped out in detail with ARPES, is the SS formed at the center of the 2d Brillouin zone (BZ) on the Be (0001) surface. That SS band is parabolic and has a measured effective mass of about 1.5 m_e/76/. LDA calculations on 1, 2, and 3 layer films of Be (0001)/56/ reveal an energy band which appears to be that SS (or its precursor), in qualitative agreement with experiment. However, the calculated effective mass of that band is only 1.0 m_e for all three film thicknesses considered. It has been suggested/56/ that this discrepancy is due to the relatively localized nature of that SS; the argument being that for a localized surface state, the self-interaction problems known to occur in localized states in other systems/51/ will cause the SS to extend too far into the vacuum region, thereby reducing the effective mass of the SS. If this analysis is correct, it could be indicative of fundamental problems associated with applying LDA to surface states in general.

The vast majority of existing LDA calculations for 2d systems have tended to focus on the one-electron properties we have discussed up to this point. Only recently have LDA film calculations achieved the high precision and speed necessary to determine the structural properties via total energy calculations. Even now, the range of problems accessible to

such calculations is limited by the complexity of the systems involved and may be divided into three categories: 1) total geometrical optimization of relatively simple films, monolayer or dilayer/62,77-82/,3) surface relaxation calculations in which the vertical spacing of the outermost layers for a thick film are allowed to relax/83,84/, and 4) surface reconstruction calculations in which the general positions of the atoms in the outermost layer are varied/85-87/.

Of the three categories of structural calculations described above, the most formidable is the surface reconstruction problem. The basic difficulty is two fold. First, the size of a single calculation for a given reconstruction will generally be quite large, with an n X m reconstruction requiring a least nm atoms per layer. Second, even when the general form or symmetry of a reconstruction is known the actual atomic positions may be unclear. Thus, the theorist can only try to sample the more obvious possible reconstructions that are consistent with the underlying symmetry and minimize the energy (see Ref. 87).

REFERENCES

1. P. Strange, J. Stauton, and B. L. Gyorffy (1984). J. Phys. C17, 3335. G. Schadler, P. Weinberger, A. M. Boring, and R. C. Albers, Phys. Rev. B34, 713.
2. V. L. Moruzzi, J. F. Janak, and A. R. Williams (1978). In "Calculated Electronic Properties of Metals" (Pergamon, New York).
3. M. S. S. Brooks (1985). In "Structure and Bonding 59/60". (L. Manes, eds.), p. 263, Springer Verlag, Berlin. M. S. Brooks, D. Johansson, and H. L. Skriver (1984). In "Handbook on the Physics and Chemistry of the Actinides". (A. J. Freeman and G. H. Lander, eds.), p. 153, Elsevier, New York.
4. J. R. Smith, J. Fenante, and J. H. Rose (1982). Phys. Rev. B25, 1419.
5. A. R. Williams and U. von Barth (1983). In "Theory of the Inhomogenous Electron Gas" (S. Lundqvist and N. H. March, eds.), p. 189, Academic Press, New York.
6. C. Kittel (1966). In "Introduction to Solid State Physics" (third Edition), p. 111, Wiley and Sons, Inc.
7. N. E. Christensen (1984), Solid State Commun. 49, 701.
8. S. P. Chen, A. F. Voter, R. C. Albers, A. M. Boring, and P. J. Hay (1989). Scripta Metall. 23, 217.
9. N. E. Christensen (1984). Ψ_k Newsletter 8, 2. N. E. Christensen (1984). Solid State Comm. 49, 701.
10. J. M. Wills (1989). Private Communication.
11. J. Chen, L. L. Bayer, H. Krakauer, and M. J. Mehel (1988). Phys. Rev. B37, 3295.
12. H. L. Skriver (1985). Phys. Rev. B31, 1909.
13. D. Glötzel and A. K. McMahan (1979). Phys. Rev. B20, 3210.
14. J. A. Moriarty and A. K. McMahan (1982). Phys. Rev. Lett. 48, 809.
15. A. K. McMahan and J. A. Moriarty (1983). Phys. Rev. B27, 3235. A. K. McMahan (1986). Physica 139, 31; and refs. therein.
16. H. Olijnyk and W. B. Holzapfel (1985). Phys. Rev. B31, 4682.
17. J. C. Boettger and R. C. Albers (1989). Phys. Rev. B39, 3010.
18. H. J. F. Jansen and S. S. Peng (1988). Phys. Rev. B37, 2689.
19. C. S. Wang, B. M. Klein, and H. Krakauer (1985). Phys. Rev. Lett. 54, 1852.
20. K. B. Hathaway, H. J. F. Jansen, and A. J. Freeman (1985). Phys. Rev. B31, 7603.
21. M. Cordona and L. Ley (1978). In "Photoemission in Solids I" (Cordona and Ley, eds.), p. 1, Springer-Verlag.
22. C. Caroli, D. Lederer-Rozenblatt, B. Roulet, and D. Saint-James (1973). Phys. Rev. B8, 4552.
23. P. J. Feibelman and D. E. Eastman (1974). Phys. Rev. B10, 4932.

24. H. Winter, P. J. Durham, and G. M. Stocks (1984). J. Phys. F14, 1097.
25. P. Marksteiner, P. Weinberger, R. C. Albers, A. M. Boring, and G. Schadler (1986). Phys. Rev B34, 6730.
26. A. M. Boring, R. C. Albers, G. Schadler, P. Marksteiner, and P. Weinberger (1987). Phys. Rev. B35, 2447.
27. F. Wooten (1980). In "Optical Properties of Solids" (Academic, New York). M. L. Cohen and J. R. Chelikowsky (1988). In "Electronic Structure and Optical Properties of Semiconductors" (Springer-Verlag, Berlin).
28. N. W. Ashcroft and K. Sturm (1970). Phys. Rev. B3, 1898.
29. M. Alouani, J. M. Kock, and M. A. Khan (1986), J. Phys. F16, 473.
30. A. R. Mackintosh and O. K. Andersen (1980). In "Electrons at the Fermi Surface," M. Springford (Cambridge), pp. 149-224.
31. N. E. Christensen (1984). Phys. Rev. B30, 5753. M. Alouani, S. Gopalan, M. Garriga, and N. E. Christensen (1988). Phys. Lett. 61, 1643.
32. J. E. Müller and J. W. Wilkins (1984). Phys. Rev. Lett. 50, 267.
33. L. A. Grunes (1983). Phys. Rev. B27, 2111.
34. G. Materlik, and J. E. Müller, and J. W. Wilkins (1983). Phys. Rev. Lett 50, 267.
35. J. J. Rehr, private communication.
36. "Electrons at the Fermi Surface" (1980), edited by M. Springford (Cambridge).
37. G. G. Lonzarich (1980). In "Electrons at the Fermi Surface," M. Springford (Cambridge), pp. 225-277.
38. K. Schoenhammer and O. Gunnarson (1988). Phys. Rev. B37, 3128.
39. R. Rasolt and S. H. Vosko (1974). Phys. Rev. B10, 4195. S. B. Nickerson and S. H. Vosko (1976). Phys. Rev. B14, 4399.
40. C. S. Wang, M. R. Norman, R. C. Albers, A. M. Boring, W. E. Pickett, H. Krakauer, and N. E. Christensen (1987). Phys. Rev. B35, 7260. M. R. Norman, R. C. Albers, A. M. Boring, and N. E. Christensen (1988). Solid State Communications 68, 245.
41. N. E. Christensen, O. K. Andersen, O. Gunnarson, and O. Jepsen (1988). J. Magn. Magn. Mater. 76 & 77, 23.
42. J. W. Wilkins (1980). In "Electrons at the Fermi Surface," edited by M. Springford (Cambridge), pp. 47-101.
43. G. Grimvall (1981). In "The Electron-Phonon Interaction in Metals, (North Holland, Amsterdam).
44. R. C. Albers, A. M. Boring, and N. E. Christensen (1986). Phys. Rev. B33, 8116.
45. A. K. McMahan, R. M. Martin, and S. Satpathy (1988). Phys. Rev. B38, 6650. O. Gunnarsson, O. K. Andersen, O. Jepsen, and J. Zaanen (1989). Phys. Rev. B39, pp. 1708. M. Schluter, M. S. Hybertsen, and N. E. Christensen (1988). Physica C 153-155, pp. 1217.

46. G. R. Stewart, Rev. Mod. Phys. **56**, 755 (1984).
47. P. A. Lee, T. M. Rice, J. W. Serene, L. J. Sham, and J. W. Wilkins (1988). Comm. Cond. Matter Phys. **12**, 99.
48. P. Fulde, J. Keller, and G. Zwicknagl (1988). Solid State Physics **41**, 1.
49. L. Tailler and G. G. Lonzarich (1988). Phys. Rev. Lett. **60**, 1570.
50. J. W. Allen, S.-J. Oh, L. E. Cox, W. P. Ellis, M. S. Wire, Z. Fisk, J. L. Smith, and A. J. Arko (1985). Phys. Rev. Lett. **54**, 2635.
51. J. W. Allen (1988). J. Magn. Mater. **76 & 77**, 324.
52. G. Zwicknagl (1988) J. Mag. Magn. Mater. **76 & 77**, 16.
53. J. O. Porteus (1974). Surf. Sci. 41, 515.
54. X. Sun, M. Farjam, and C-W Woo (1983). Phys. Rev. B28, 5599; and refs. therein.
55. N. D. Lang and W. Kohn (1970). Phys. Rev. B1, 4555; Phys. Rev. B3, 1215 (1971).
56. J. C. Boettger and S. B. Trickey (1986). Phys. Rev. B34, 3604.
57. J. P. Perdew and A. Zunger (1981). Phys. Rev. B23, 5048. L. Fritshe and H. Gollisch (1982). Z. Phys. B-Cond. Mat. 48, 209.
58. N. V. Smith (1985). Phys. Rev. B32, 3549.
59. J. E. Inglesfield (1982). Rep. Prog. Phys. 45, 223; and refs. therein.
60. J. F. Janak 91978). Phys. Rev. B18, 7165; M. Levy, J. P. Perdew, and V. Sahni (1984). Phys. Rev. A30, 2745.
61. E. Wimmer (1983). J. Phys. F13, 2313; J. Phys. F14, 681 (1984).
62. J. C. Boettger and S. B. Trickey, J. Phys. F (accepted) .
63. F. K. Schulte (1976). Surf. Sci. 55, 427.
64. P. J. Feibelman (1983). Phys. Rev. B27, 1991.
65. P. J. Feibelman and D. R. Hamann (1984). Phys. Rev. B29, 6463.
66. I. P. Batra, S. Ciraci, G. P. Srivastava, J. S. Nelson, and G. Y. Fong (1986). Phys. Rev. B34, 8246.
67. M. Jalochowski and E. Bauer (1988). Phys. Rev. B38, 5272; and refs. therin.
68. J. Zak (1985). Phys. Rev. B32, 2218; and refs. therein.
69. E. V. Chulkov, V. M. Silkin, and E. N. Shirykalov (1987). Surf. Sci. 188, 287.
70. M. El-Batanouny, D. R. Hamann, S. R. Chubb, and J. W. Davenport, Phys. Rev. B27, 2575.
71. J. Redinger, C. L. Fu, A. J. Freeman, U. Konig, and P. Weinberger (1988). Phys. Rev. B38, 5208.
72. E. Wimmer, A. Neckel, and A. J. Freeman (1985). Phys. Rev. B31, 2370.
73. G. W. Fernando and B. R. Cooper (1988). Phys. Rev. B38, 3016.
74. M. Posternak, H. Krakauer, A. J. Freeman, and D. D. Koelling (1980). Phys. Rev. B21, 5601.

75. B. Cord, R. Courths, and H. Wern (1985). Phys. Rev. B31, 1164.
76. R. A. Bartynski, E. Jensen, T. Gustafsson, and E. W. Plummer (1985). Phys. Rev. B32, 1921.
77. H. J. F. Jansen, A. J. Freeman, M. Weinert, and E. Wimmer (1983). Phys. Rev. B28, 593.
78. E. Wimmer (1983) Surf. Sci. 134, L487.
79. J. C. Boettger and S. B. Trickey (1984). J. Phys. F14, L151.
80. I. P. Batra, J. Vac. Sci. Techn. A3, 1603 (1985).
81. J. C. Boettger and S. B. Trickey (1985). Phys. Rev. B32, 1356.
82. J. C. Boettger and S. B. Trickey (1986). J. Phys. F16, 693.
83. D. R. Hamann and P. J. Feibelman (1988). Phys. Rev. B37, 3847.
84. P. J. Feibelman (1988). Phys. Rev. B38, 7287.
85. J. Ihm, M. L. Cohen, and D. J. Chadi (1980). Phys. Rev. B21, 4592.
86. S. Ciraci and I. P. Batra (1988). Phys. Rev. B37, 2955.
87. C. L. Fu and A. J. Freeman (1988). Phys. Rev. B37, 2685.

Index

D

E

F

G

H

I

J

K

L

M

N

O

T

U

V

W

X