N-Representability in the Quantum Kernel Energy Method

Walter Polkosnik

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Quantum Physics Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/2610
$N$-Representability in the Quantum Kernel Energy Method

by

Walter Polkosnik

A dissertation submitted to the Graduate Faculty in Physics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2018
This manuscript has been read and accepted by the Graduate Faculty in Physics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Professor Louis J. Massa

Date  Chair of Examining Committee

Professor Igor Kuskovsky

Date  Executive Officer

Professor Leon Cohen
Professor Steve G. Greenbaum
Professor John Lombardi
Professor Chérif F. Matta

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

*N*-REPRESENTABILITY IN THE QUANTUM KERNEL ENERGY METHOD

by

WALTER POLKOSNIK

Adviser: Professor Louis J. Massa

The Kernel Energy Method (KEM) delivers accurate full molecule energies using less computational resources than standard *ab-initio* quantum chemical approaches. KEM achieves this efficiency by decomposing a system of atoms into disjoint subsets called *kernels*. The results of full *ab-initio* calculations on each individual single kernel and on each double kernel formed by the union of each pair of single kernels are combined in an equation of a form that is specific to KEM to provide an approximation to the full molecule energy. KEM has been demonstrated to give accurate molecular energies over a wide range of systems, chemical methods and basis sets. The efficiency of KEM makes calculations on large systems tractable. It has been shown to be accurate for calculations on large biomolecules including proteins and DNA.

KEM has recently been generalized to provide the one-body density matrix for the full molecule. This generalization greatly expands the utility of the method since the density matrix enables simple calculation of the expectation value of any observable associated with a one-body operator. More importantly, the generalization enables results from KEM to be constrained to satisfy essential quantum mechanical conditions.

KEM is generalized to density matrices by taking the density matrices obtained from the single and double kernels and adapting them to the full molecule basis. These *augmented* kernel density matrices are summed according to the standard KEM expansion. This initial matrix does not necessarily correspond to a valid quantum mechanical *N*-electron wavefunction which is normalized and obeys the Pauli principle. Such a density matrix is called *N*-representable. *N*-representability
is imposed on the initial matrix by using the Clinton equations, which deliver a single-determinant $N$-representable one-body density matrix. Single-determinant $N$-representability is ensured by enforcing the density matrix to be a normalized projector. Such normalized projectors are factorizable into matrices which deliver full molecule molecular orbitals.

Although energies have been demonstrated to be accurate in the original energy expansion form of KEM, there is no requirement that the energy corresponds to the expectation value of a valid quantum mechanical wavefunction. Because of this, there is no requirement that the energies satisfy the variational theorem. Imposing single-determinant $N$-representability on the KEM density matrix by using the Clinton equations recovers the variational bound on the energy obtained from KEM.

Recovery of the variational bound by the $N$-representable KEM density matrix has been demonstrated by calculations on several water clusters. Energies from full molecule restricted Hartree-Fock calculations were compared to energies obtained from the KEM energy expansion and to energies obtained from the KEM density matrix expansion which had $N$-representability imposed. In each case where the energy expansion gave energies that violated the variational bound the $N$-representable density matrix gave energies that satisfied the bound and in fact were more accurate than the simple energy expansion results.

The effect of imposing $N$-representability on the KEM density matrix has also been investigated in the context of Kohn-Sham Density Functional Theory (KS/DFT). Calculations on a simple noble gas system and a single water cluster were done. KS/DFT energies were compared to energies from the KEM energy expansion and energies associated with the KEM density matrix expansion on which $N$-representability was imposed. The energies from the KEM density matrix expansion made $N$-representable were more accurate than those obtained from straightforward KEM energy expansions in nearly all cases for the noble gas system. For the water cluster the accuracy of the energy obtained from the $N$-representable KEM density matrix was nearly four times closer to the energy calculated for the full system than the KEM energy expansion result.
The Clinton-purified \(N\)-representable matrix obtained from the KEM density expansion can be used to calculate expectation values for any one-electron operator. In particular, a KEM density matrix approach can be used in the study of quantum crystallography.
Dedication

Dedicated to Prof. Narciso Garcia
Acknowledgments

I thank all the members of my committee: Leon Cohen, Steve Greenbaum, John Lombardi and Chérif Matta. I am particularly grateful to Steve Greenbaum as he was the executive officer when I re-entered the Physics program.

I cannot imagine a better advisor than Lou Massa. It was an honor and a privilege to work with him.

Daniel Moy in the Physics department was exceedingly helpful throughout my entire time in the program. I also am grateful to the current executive officer of the Physics program, Igor Kuskovsky, for all his help.

I am grateful to the Journal of Computational Chemistry and Acta Physico-Chimica Sinca for permission to reprint my publications.

I thank my friend Francisco Iovino for his exuberant encouragement.

I especially want to thank Deena Patel for inspiring me to ask the important questions, both in physics and in life.
# Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 The $N$-electron Schrödinger Equation .......... 2
1.2 The Variational Theorem ........................................ 9
1.3 Hartree-Fock Theory ........................................ 11
1.4 Configuration Interaction .............................. 15
1.5 Density Functional Theory ............................ 17

## 2 $N$-representability

2.1 Density matrices ........................................ 22
2.2 Necessary and sufficient conditions for $N$-representability ........................................ 25

## 3 Clinton Equations

3.1 The Clinton equations for general constraints ....... 30
3.2 Single-determinant $N$-representability via the Clinton equations .............................. 32
4 X-ray Crystallography
  4.1 X-ray diffraction .............................. 34
  4.2 Structure determination ...................... 36
  4.3 Quantum Crystallography ...................... 36

5 The Kernel Energy Method ......................... 39
  5.1 The KEM equations ............................ 39
  5.2 KEM and N-representability ................... 41

6 KEM Hartree-Fock calculations on Water Clusters 45

7 KEM KS/DFT calculations on Noble Gas and Water Cluster 50

8 Discussion and Conclusions ....................... 57

Appendices ........................................... 60

A Basis Functions .................................... 61

B Water Cluster publication ........................ 63

C KS/DFT publication ................................ 70

D Submitted Quantum Crystallography publication 77

Bibliography .......................................... 93
List of Tables

1.1 Atomic units ......................................................... 4

6.1 KEM energy and energy of $N$-representable density matrix compared to full molecule Hartree-Fock energy across water clusters. ........................................ 48

7.1 KEM energy compared to full molecule KS/DFT energy for a cluster of twelve water molecules .................................................. 51

7.2 Energy of $N$-representable density matrix compared to full molecule KS/DFT energy for a cluster of twelve water molecules .................................................. 52

7.3 KEM energy and energy of $N$-representable density matrix compared to full molecule KS/DFT energy for the linear helium system at various separations, STO-3G basis .......................................................... 53

7.4 KEM energy and energy of $N$-representable density matrix compared to full molecule KS/DFT energy for the linear helium system at various separations, 6-31G($d$, $p$) basis .......................................................... 55
List of Figures

1.1 Vectors specifying a molecular system ........................................... 2

2.1 Mapping of $N$-representable density matrices and non-$N$-representable functions . 28

4.1 X-ray scattering geometry ............................................................... 35

6.1 $N = 15$ water cluster, fused pentameric structure ............................... 47
6.2 $N = 20$ water cluster, fused tetrameric structure ............................... 47
6.3 KEM energy and energy of $N$-representable density matrix compared to full molecule Hartree-Fock energy across water clusters. ........................................ 49

7.1 KEM energy and energy of $N$-representable density matrix compared to full molecule KS/DFT energy for the linear helium system at various separations in the STO-3G basis. ................................................................. 54
7.2 KEM energy and energy of $N$-representable density matrix compared to full molecule KS/DFT energy the linear helium system at various separations in the 6-31G($d, p$) basis. ................................................................. 56
Understanding the quantum mechanics of electrons in molecules is fundamental to understanding chemistry.\textsuperscript{1–3}

The computational difficulty of the most commonly used \textit{ab-initio} approaches to solving the fundamental quantum mechanical equations scale polynomially in the size of the system. This is a barrier to understanding large systems. The slowest scaling theories such as Hartree-Fock Theory and Density Functional Theory scale as $N^4$, where $N$ is the number of atoms in the system. More accurate methods scale at higher powers of $N$. These methods become unfeasible quickly as the number of atoms in the system increases. Some of the molecules of greatest interest, biomolecules such as proteins and DNA, consist tens of thousands or even hundreds of thousands of atoms.
Figure 1.1: Vectors specifying a molecular system. Vectors \( \mathbf{R}_A, \mathbf{R}_B \) specify nuclei. Vectors \( \mathbf{r}_i, \mathbf{r}_j \) specify electrons.

### 1.1 The \( N \)-electron Schrödinger Equation

The time-independent Hamiltonian for a non-relativistic system of \( N \) electrons and \( M \) nuclei is

\[
\hat{H} = \sum_{A=1}^{M} \left( -\frac{1}{2M_A} \nabla_A^2 \right) + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{|\mathbf{R}_B - \mathbf{R}_A|} + \\
\sum_{i=1}^{N} \left( -\frac{1}{2} \nabla_i^2 \right) + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{-Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} \tag{1.1}
\]

where lowercase symbols refer to electrons and uppercase symbols refer to nuclei. The position vector of electron \( i \) is \( \mathbf{r}_i = (x_i, y_i, z_i) \) and \( \nabla_i^2 \) is the Laplacian with respect to those coordinates,
\[ \nabla_i^2 = \left( \frac{\partial^2}{\partial x_i^2}, \frac{\partial^2}{\partial y_i^2}, \frac{\partial^2}{\partial z_i^2} \right) \text{ with } i = 1 \ldots N. \]

Similarly for the nuclei, the position vector of nucleus \( A \) is \( \mathbf{R}_A \) and \( \nabla_A^2 \) is the Laplacian associated with the position of nucleus \( A \), \( \nabla_A^2 = \left( \frac{\partial^2}{\partial x_A^2}, \frac{\partial^2}{\partial y_A^2}, \frac{\partial^2}{\partial z_A^2} \right) \), with \( A = 1 \ldots M \). The mass of nucleus \( A \) is \( M_A \) and its atomic number is \( Z_A \).

The first two terms in equation 1.1 refer only to nuclear coordinates. The first term, \( \hat{T}_N \), is the sum of kinetic energy operators for the nuclei. The second term is the total nuclear-nuclear potential energy, \( \hat{V}_N \).

\[ \hat{H}_N = \hat{T}_N + \hat{V}_N = \sum_{A=1}^{M} \left( -\frac{1}{2M_A} \nabla_A^2 \right) + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{|\mathbf{R}_B - \mathbf{R}_A|} \quad (1.2) \]

The remaining terms make up the electron Hamiltonian. The first term on the second line of equation 1.1 is the electronic kinetic energy, \( \hat{T}_e \). The second term is the electron-electron potential energy, \( \hat{V}_{ee} \). The last term is the potential energy due to electron-nuclear interactions, \( \hat{V}_{Ne} \).

\[ \hat{H}_e = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{Ne} = \sum_{i=1}^{N} \left( -\frac{1}{2} \nabla_i^2 \right) + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{-Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} \quad (1.3) \]

Each electron has an additional internal degree of freedom, spin. Electron spin does not appear in the Hamiltonian of equation 1.1 but plays an important role in obtaining solutions which satisfy the Pauli principle.

Equation 1.1 is expressed units convenient for consideration of the electronic part of the Hamiltonian. Mass and charge are measured relative to the electronic mass and charge. Distance and energy are measured relative to the radius, \( a_0 \), and energy, \( E_0 \), of the electron in the ground state of the Bohr model of the hydrogen atom. These atomic units (a.u.) are summarized in table 1.1.

The Born-Oppenheimer approximation

For the least massive atomic nucleus, that of hydrogen, \( m_e/m_{\text{proton}} \approx 5 \times 10^{-4} \). The Born-Oppenheimer approximation was originally developed as a time-independent perturbation expansion of the Hamil-
### Table 1.1: Atomic units. The definition of each dimension is given in terms of its expression in SI units.

- **Mass**: $m_e$ a.u.
- **Charge**: $e$ a.u.
- **Length**: $a_0 = \frac{4\pi\varepsilon_0\hbar}{m_e e^2}$ Bohr
- **Energy**: $E_h = \frac{\hbar^2}{m_e a_0^2} = 2E_0$ Hartree

atomic in equation 1.1 with the nuclear kinetic energy taken as a small perturbation relative to the remaining terms in the Hamiltonian along with a Taylor expansion in the displacements of the nuclei, which are assumed to be confined near a set of fixed positions $\bar{R}_A$.\textsuperscript{5,6} The perturbation expansion is in powers of $\kappa = (m_e/M)^{1/4}$, where $M$ is taken as any of the nuclear masses or their mean. The nuclei are considered as fluctuating near some particular configuration $\bar{R}_A$, $A = 1\ldots M$ with their displacement defined as $\kappa U_A = R_A - \bar{R}_A$.

The total Hamiltonian is written

$$\hat{H} = \hat{H}_0 + \kappa^4 \hat{T}_N$$ \hspace{1cm} (1.4)

where $\hat{T}_N$ is the nuclear kinetic energy term, considered a small perturbation to $\hat{H}_0$, which contains the remaining terms of the full Hamiltonian of equation 1.1.

The solutions for $\hat{H}_0$, which involve only the electronic motion with the nuclei at fixed positions are given by

$$\hat{H}_0 \Psi(r_1\ldots r_N; R_1\ldots R_M) = E_0(R_1\ldots R_M)\Psi(r_1\ldots r_N; R_1\ldots R_M)$$ \hspace{1cm} (1.5)

The solutions to this equation are assumed to be known for the configuration $R_A = \bar{R}_A$ and for all nearby configurations. Both $E_0$ and $\Psi$ depend on the nuclear coordinates parametrically; they are considered constants as far as the eigenvalue equation is concerned.
Expansion to the first few orders in both nuclear displacement $U_A$ and in the perturbation expansion leads to three important results. The first is that

$$\nabla_B E(R_1 \ldots R_M)_{R_A} = 0, \quad A, B = 1 \ldots M \quad (1.6)$$

In other words, the nuclei fluctuate about an equilibrium configuration determined by the stationary point of the energy.

The second is that the nuclear and electronic wavefunctions factorize to a product of a nuclear wavefunction, $\phi$, that depends only on nuclear coordinates, and an electronic wavefunction, $\psi$, that depends on the nuclear positions only parametrically. The wavefunction can be written

$$\Psi(r_1 \ldots r_N; U_1 \ldots U_M) = \phi(U_1 \ldots U_M) \psi(r_1 \ldots r_N; R_1 \ldots R_M) \quad (1.7)$$

The first factor describes the nuclear motion as a function of the displacement of the nuclei from their equilibrium positions. The second factor describes the electronic motion, which depends parametrically on the instantaneous positions of the nuclei.

The picture of a molecular system to low orders in the Born-Oppenheimer approximation is that of dynamic electrons moving in the field of the nuclei. The electronic wavefunction reacts instantaneously to the motion of the nuclei. In this motion the electronic wavefunction is deformed smoothly as a function of the nuclear coordinates; electrons do not make transitions from one state to another. The results up to this order of approximation are therefore called the adiabatic approximation.

The third result is that the effective potential for the nuclei is a sum of terms of the second, third and forth power of the nuclear displacements $U_1 \ldots U_M$. The nuclei move in a harmonic potential perturbed by third and fourth order anharmonic terms. The anharmonic terms are of order $\kappa$ and $\kappa^2$ respectively, relative to the order of the harmonic term.

Taking into account overall translational and rotational motion of the molecule, the perturba-
tion series leads to an equation that includes the coupling between the rotation, oscillation and electronic motion of the molecule and the overall rotation and translation of the molecule.

The factorization of the wavefunction for the full system does not continue at higher orders of the perturbation expansion. The details of the coupling between the electronic and nuclear motion can be more clearly seen by starting with the assumption that the wavefunction can be written as a product.\cite{6,7}

Writing the full $N$-electron Hamiltonian in terms of the displacement of the nuclei from fixed positions $U_A = R_A - \bar{R}_A$,

$$
\hat{H} = \sum_{A=1}^{M} \left(-\frac{1}{2M_A} \nabla_A^2\right) + \sum_{i=1}^{N} \left(-\frac{1}{2} \nabla_i^2\right) + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|r_i - r_j|} + \hat{V}_{Ne}(r_1 \ldots r_N; U_1 \ldots U_M) + \hat{V}_N(U_1 \ldots U_M)
$$

(1.8)

where $\hat{V}_{Ne}$ represents the potential energy of electron-nuclear interactions and $\hat{V}_N$ represents the nuclear-nuclear potential energy. The Laplacian, $\nabla_A$, is taken with respect to the nuclear displacements.

Assume

$$
\Psi(r_1 \ldots r_N; U_1 \ldots U_M) = \phi(U_1 \ldots U_M) \psi(r_1 \ldots r_N; U_1 \ldots U_M)
$$

(1.9)

with $\psi$ satisfying the electronic part of the Hamiltonian in field of static nuclei,

$$
\hat{H}_e \psi(r_1 \ldots r_N; U_1 \ldots U_M) = E_e(U_1 \ldots U_M) \psi(r_1 \ldots r_N; U_1 \ldots U_M)
$$

(1.10)

where

$$
\hat{H}_e = \sum_{i=1}^{N} \left(-\frac{1}{2} \nabla_i^2\right) + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|r_i - r_j|} + \hat{V}_{Ne}(r_1 \ldots r_N; U_1 \ldots U_M)
$$

(1.11)

and the eigenvalues $E_e$ are a function of $U$. 
Applying the full Hamiltonian to the product wavefunction,

\[ \hat{H}\Psi = \psi \left[ \sum_A \frac{-1}{2M_A} \nabla_A^2 + V_N(U_1 \ldots U_M) + E_e(U_1 \ldots U_M) \right] \phi + \sum_A \frac{-1}{2M_A} \left[ 2\nabla_A \phi \cdot \nabla_A \psi + \phi \nabla_A^2 \psi \right] \] (1.12)

The first line can be interpreted as a Schrödinger equation for the nuclei alone, along with an additional potential, \( E_e(U_1 \ldots U_M) \), which is the energy of the electronic Hamiltonian as a function of the position of the nuclei.

\[ \left[ \sum_A \frac{-1}{2M_A} \nabla_A^2 + E_e(U_1 \ldots U_M) + V_N(U_1 \ldots U_M) \right] \phi = E\phi \] (1.13)

This energy accounts for the interaction between the nuclei and electrons. The nuclei respond to the changes in the electronic energy as the electrons respond adiabatically to the motion of the nuclei.

The remaining non-adiabatic terms can be treated by time independent perturbation theory. The first term does not contribute at the zeroth order of perturbation theory as the matrix elements contain factors in the integrand of the form

\[ \int dr \psi^* \nabla_A \psi = \frac{1}{2} \nabla_A \int dr \psi^* \psi = \nabla_A 1 = 0 \] (1.14)

since, for stationary states, the \( \psi \) can be taken as real. The off diagonal terms contribute at second order of perturbation theory and give amplitudes for transitions between electronic states as the nuclei move. This is termed the *electron-phonon* interaction.

The expectation value of the second term contains factors in the full integrand of the form

\[ -\left( \frac{1}{2M_A} \right) \int dr \psi^* \nabla_A^2 \psi = \frac{m_e}{M_A} \int dr \left( -\frac{1}{2m_e} \right) \psi^* \nabla_A^2 \psi \] (1.15)
where the electron mass has been restored in the equations for clarity. The largest value of these contributions will be when the electron $i$ is tightly bound to a nucleus $A$, in which case $\psi (r_i; U_A) = \psi (r_i - U_A)$ and

$$\frac{m_e}{M_A} \int d\mathbf{r} \left( -\frac{1}{2m_e} \right) \psi^\ast \nabla^2_A \psi = \frac{m_e}{M_A} \int d\mathbf{r} \left( -\frac{1}{2m_e} \right) \psi^\ast \nabla_i^2 \psi$$

(1.16)

This contribution is then $m_e/M \lesssim 5 \times 10^{-4}$ times the expectation value of the kinetic energy of the electrons and can be neglected.

The Born-Oppenheimer approximation, apart from higher order effects, allows a factorization of the total wavefunction into a product of a nuclear wavefunction and an electronic wavefunction. The electrons move as if they are in a field of fixed nuclei while the nuclei move on a potential energy hypersurface determined by the nuclear-nuclear potential energy and the electronic energy.

For a modern overview of the Born-Oppenheimer approximation see Sutcliffe. For a mathematical treatment see Jecko et. al.

The electronic Schrödinger equation

The remainder of this thesis will focus on the electronic Hamiltonian. The nuclei will be assumed to be at fixed positions and the functional dependence of the electronic energy and electronic wavefunctions on nuclear coordinates will be omitted from equations.

The time independent Schrödinger eigenvalue equation for the electrons in the Born-Oppenheimer approximation is

$$\hat{H}\Psi (x_1, \ldots, x_N) = E\Psi (x_1, \ldots, x_N)$$

(1.17)

$$\hat{H} = \sum_{i=1}^{N} \left( -\frac{1}{2} \nabla_i^2 \right) + \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{-Z_A}{|r_i - R_A|} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|r_i - r_j|}$$

(1.18)

Electrons have an additional fundamental internal degree of freedom, spin. While spin does
CHAPTER 1. INTRODUCTION

not appear in the Hamiltonian of equation 1.18, it plays an important role in obtaining quantum mechanically valid solutions to the eigenvalue equation, equation 1.17. The spin degrees of freedom are included symbolically through the use of variables $x_i$ in equation 1.17, which represent the position of each electron $r_i$ along with its spin, denoted in this shorthand as $\sigma_i$, so that $x_i = (r_i, \sigma_i)$.

A quantum mechanically valid wavefunction must be antisymmetric in exchange of any two electrons.

$$\Psi(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_N) = -\Psi(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_N)$$  \hspace{1cm} (1.19)

This mathematical constraint enforces the Pauli exclusion principle, since if any two electrons have the same position and spin,

$$\Psi(x_1, \ldots, x_i, \ldots, x_i, \ldots, x_N) = -\Psi(x_1, \ldots, x_i, \ldots, x_i, \ldots, x_N) = 0$$  \hspace{1cm} (1.20)

and so there is zero probability of finding two electrons at identical positions with identical spin states.

1.2 The Variational Theorem

The time-independent eigenvalue equation for a given Hamiltonian is

$$\hat{H}|\Psi\rangle = E|\Psi\rangle$$  \hspace{1cm} (1.21)

In the variational approach a proposed trial ket, $|\Psi'\rangle$, which is normalized and satisfies the same boundary conditions as $|\Psi\rangle$ is introduced. The trial ket generally depends on a set of parameters to be varied in such a way as to increase the accuracy of the approximation. The variational theorem guarantees that the expectation value of the Hamiltonian evaluated in any trial ket is an upper bound to the exact ground state energy.
The exact solutions to equation 1.21 will consist of a set, assumed to be discrete, of orthonormal
eigenkets $|\Psi_{ni}\rangle$ and their corresponding real eigenvalues $E_{ni}$ where \{ni\} consists of a set of
integers which label the eigenvalue and ket. The eigenvalues are real and so can be ordered so that

$$E_0 \leq E_1 \leq \cdots \leq E_n \leq \cdots$$

where without loss of generality we have introduced a single subscript which corresponds to a
particular set \{ni\} and serves to define the ordering.

The eigenkets are assumed to form a complete set

$$1 = \sum_n |\Psi_n\rangle \langle \Psi_n|$$

so the expectation value of the trial ket can be written

$$\langle \Psi'| \hat{H} | \Psi' \rangle = \sum_{mn} \langle \Psi'| \Psi_n\rangle \langle \Psi_n| \hat{H} | \Psi_m\rangle \langle \Psi_m| \Psi' \rangle = \sum_n E_n \| \langle \Psi_n| \Psi' \rangle \|^2$$

and noting that since $E_0 \leq E_n$ for all $n > 0$

$$\langle \Psi'| \hat{H} | \Psi' \rangle \geq E_0 \sum_n \| \langle \Psi_n| \Psi' \rangle \|^2$$

and

$$E_0 \sum_n \| \langle \Psi_n| \Psi' \rangle \|^2 = E_0$$

then

$$\langle \Psi'| \hat{H} | \Psi' \rangle \geq E_0$$

with equality satisfied if and only if $|\Psi\rangle = |\Psi_0\rangle$, a ground state of the system.

The main result is that the expectation value of any trial ket gives an upper bound to the ground
state energy of a system.

The quality of the approximation is measured by the value of the expectation value of the energy. Trial kets which have lower expectation values are better approximations to the ground state energy. For trial kets that depend on a set parameters, the parameters which lead to the lowest expectation value give the best approximation to the ground state energy for that set of trial functions.

The variational theorem holds for any mathematically acceptable trial ket. There are approaches to approximate the ground state energy of a system that do not use state kets, and in these approaches the variational bound does not necessarily hold.

1.3 Hartree-Fock Theory

In the Hartree-Fock approximation the \( N \)-electron wavefunction is taken as a single Slater determinant of exactly \( N \) one-electron spin orbitals.\(^{11}\) The spin orbitals are formed of products of a spatial orbital and a spin ket, \( \psi_j(x_i) = \phi_j(r_i) |\sigma_i\rangle \) with \( |\sigma_i\rangle \) taken as spin up or spin down. The spatial orbitals are taken to be orthonormal.

\[
\Psi_{\text{HF}} = \frac{1}{\sqrt{N!}} \begin{vmatrix}
\psi_i(x_1) & \psi_j(x_2) & \cdots & \psi_k(x_1) \\
\psi_i(x_2) & \psi_j(x_2) & \cdots & \psi_k(x_2) \\
\vdots & \vdots & \ddots & \vdots \\
\psi_i(x_N) & \psi_j(x_N) & \cdots & \psi_k(x_N)
\end{vmatrix} = |\psi_1 \cdots \psi_N\rangle
\] (1.28)

The Slater determinant 1.28 is manifestly antisymmetric with respect to interchange of electrons and is normalized due to the orthonormality of the spin orbitals.
CHAPTER 1. INTRODUCTION

The energy expectation value can be written explicitly in terms of the individual spin orbitals,

\[ E_{\text{HF}} = \sum_{i}^{N} H_i + \frac{1}{2} \sum_{i,j=1}^{N} (J_{ij} - K_{ij}) \]  

(1.29)

with

\[ H_i = \int d\mathbf{x} \psi_i^*(\mathbf{x}) \left[ \left( -\frac{1}{2} \nabla^2 \right) + V_{\text{Ne}}(\mathbf{x}) \right] \psi_i(\mathbf{x}) \]  

(1.30)

\[ J_{ij} = \int d\mathbf{x}_1 d\mathbf{x}_2 \psi_i^*(\mathbf{x}_1) \psi_j(\mathbf{x}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_j^*(\mathbf{x}_2) \psi_j(\mathbf{x}_2) \]  

(1.31)

\[ K_{ij} = \int d\mathbf{x}_1 d\mathbf{x}_2 \psi_i^*(\mathbf{x}_1) \psi_j(\mathbf{x}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_i^*(\mathbf{x}_2) \psi_i(\mathbf{x}_2) \]  

(1.32)

The \( J_{ij} \) are called the Coulomb integrals and the \( K_{ij} \) are called the exchange integrals.

Varying the above equations with orthonormality enforced as a constraint leads to the Hartree-Fock integro-differential equations,

\[ \hat{F} \psi_i(\mathbf{x}) = \sum_{j}^{N} \epsilon_{ij} \psi_j(\mathbf{x}) \]  

(1.33)

where

\[ \hat{F} = -\frac{1}{2} \nabla^2 + V_{\text{Ne}}(\mathbf{x}) + \hat{j} - \hat{k} \]

\[ \hat{j}(\mathbf{x}_1) u(\mathbf{x}_1) = \sum_{i=1}^{N} \int d\mathbf{x}_2 \psi_i^*(\mathbf{x}_2) \psi_i(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} u(\mathbf{x}_1) \]  

(1.34)

\[ \hat{k}(\mathbf{x}_1) u(\mathbf{x}_1) = \sum_{i=1}^{N} \int d\mathbf{x}_2 \psi_i^*(\mathbf{x}_2) \psi_i(\mathbf{x}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} u(\mathbf{x}_2) \]

with the matrix \( \epsilon_{ij} \) containing the Lagrange multipliers enforcing the orthonormalization constraint.
Roothaan equations

To solve equation 1.33 Roothaan\textsuperscript{12} expanded the molecular orbitals in the determinant 1.28 as a linear combination of a set of fixed \textit{basis functions} $\chi_\mu$, $\mu = 1 \ldots K$. This converts the integro-differential equation to a set of algebraic equations for the expansion coefficients.

Using Greek suffixes to refer to the fixed basis functions and Latin suffixes to refer to molecular orbitals, the expansion can be written

$$\phi_i = \sum_\mu C_{\mu i} \chi_\mu$$

(1.35)

In the \textit{restricted Hartree-Fock} (RHF) approximation the spatial orbitals for the spin up and spin down electrons are taken to be identical. Each orbital is taken to be doubly occupied, forming a \textit{closed shell}. Most molecules have a closed shell structure in the ground state.\textsuperscript{12} The equations for $C_{\mu i}$ for closed shell orbitals are\textsuperscript{13}

$$\mathbf{F} \mathbf{C} = \mathbf{S} \mathbf{C} \mathbf{E}$$

(1.36)

with

$$E_{ij} = e_i \delta_{ij}$$

(1.37)

$$S_{\mu \nu} = \int dr \, \chi_\mu(r) \chi_\nu(r)$$

(1.38)

$$H_{\mu \nu} = \int dr \, \chi_\mu(r) H \chi_\nu(r)$$

(1.39)

$$(\mu \nu | \lambda \sigma) = \int dr dr' \, \chi_\mu(r) \chi_\nu(r) \frac{1}{|r - r'|} \chi_\lambda(r') \chi_\sigma(r')$$

(1.40)
\[ R_{\mu\nu} = 2 \sum_{i} C_{\mu i} C_{\nu i} \]  

(1.41)

\[ F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma} R_{\lambda\sigma} \left[ (\mu\nu|\lambda\sigma) - \frac{1}{2} (\mu\lambda|\nu\sigma) \right] \]  

(1.42)

\( F \) is called the Fock matrix. The \( H \) matrix is called the core Hamiltonian and consists of a Hamiltonian for a single electron moving in the field of the nuclei. The \((\mu\nu|\lambda\sigma)\) are called the two-electron integrals and depend only on the fixed basis functions. The basis functions need not be orthonormal. This is reflected in the overlap matrix, \( S \). The density matrix in the non-orthogonal basis is denoted by \( R \).

The \( \epsilon_i \) are the Fock energies. The \( N/2 \) molecular orbitals corresponding to the lowest eigenvalues \( \epsilon_i \), are called the occupied molecular orbitals and are taken as the molecular orbitals in the Hartree-Fock Slater determinant. The remaining orbitals are called the unoccupied orbitals.

The basis functions are usually chosen to be centered on the nuclei and normally depend only on the atomic number of each nucleus. They can be chosen to be the atomic orbitals of the corresponding atom, in which case expansion 1.35 is called a linear combination of atomic orbitals. Appendix A discusses commonly used basis sets.

Equation 1.36 is nonlinear as the Fock matrix itself depends on the coefficients. It is solved using a fixed point iteration procedure called the self-consistent field (SCF) procedure. In the initial step a starting value is assumed for the coefficients of the expansion and the Fock matrix is constructed. Further iterations consist of solving the eigenvalue equation 1.36 for new coefficients and energies and updating the Fock matrix with the new coefficients. The procedure is judged to have converged when further updates lead to a change of less than a set tolerance in various measures. Measures of convergence include the root-mean-square change in the density matrix and the maximum change in the density matrix.\(^{14}\)
Scaling of Hartree-Fock methods

The computational difficulty of the Hartree-Fock method scales with the number of electrons as $\sim N^4$ with no further simplification employed. This is due to the number of two-electron integrals that need to be evaluated in equation 1.40.

1.4 Configuration Interaction

Inherent in the single Slater determinant of the Hartree-Fock model is the lack of correlation between electrons of opposite spin. Electrons of opposite spin behave as if probabilistically independent. For two electrons occupying different spatial orbitals, $\psi_1(r_1)$ and $\psi_2(r_2)$, the joint probability of finding the first electron in the vicinity of $r_1$ and simultaneously finding the second electron in the vicinity of $r_2$ is

$$P(r_1, r_2) \, dr_1 dr_2 = \frac{1}{2} \left[ |\psi_1(r_1)|^2 |\psi_2(r_2)|^2 + |\psi_1(r_2)|^2 |\psi_2(r_1)|^2 \right] \, dr_1 dr_2$$

$$= \frac{1}{2} \left[ P_1(r_1) P_2(r_2) + P_1(r_2) P_2(r_1) \right] \, dr_1 dr_2$$

(1.43)

If $\psi_1 = \psi_2$ then

$$P(r_1, r_2) = P(r_1) \cdot P(r_2)$$

(1.44)

which is the joint probability of two completely independent events.

In general the correlation energy is defined as the difference between the true energy and the limiting value of the Hartree-Fock energy as the size of the basis is increased.

$$E_{\text{correlation}} = E - E_{\text{HF}}$$

(1.45)

In order to capture the correlation between electrons of opposite spin additional determinants are included. These determinants are formed by replacing one or more of the occupied molecular orbitals in the Hartree-Fock Slater determinant by orbitals corresponding to higher Fock energies.
These are the unoccupied orbitals obtained in the Hartree-Fock procedure. Each such determinant is called a configuration or excited determinant and methods that use these additional determinants are called configuration interaction or multiconfiguration methods.

Denoting the Hartree Fock result as $|\Psi\rangle$, a singly excited determinant is formed by replacing occupied orbital $a$ by unoccupied orbital $r$, denoted by $|\Psi_r^a\rangle$. A multideterminental expansion can be written

$$|\Psi_{CI}\rangle = c_0 |\Psi\rangle + \sum_{ar} c_r^a |\Psi_r^a\rangle + \sum_{abrs} c_r^s c_s^a |\Psi_r^s^a^b\rangle + \ldots$$ (1.46)

where the coefficients $c_I$ are solved for by minimizing the energy.

If all substitutions are included the method is called full configuration interaction or FCI. FCI is computationally expensive except for small systems, so the expansion is usually truncated. If only doubly excited determinants are included the method is called CID. If both singly and doubly excited determinants are included the method is called CISD. Single determinants taken alone do not lead to lower energy. Truncating the expansion, however, leads to a loss of size consistency. When CI truncated to a given level is applied to a system that consists of distantly separated parts the total energy of the full system is not the sum of the energies of the separated parts. This is because simultaneous excitation to a certain level in the separated parts corresponds to what would necessarily be a higher order excitation in the combined system.

The coupled cluster approach is a size consistent approach that includes excited determinants. The CC expansion expansion is written as $|\Psi_{CC}\rangle = (1 + \tilde{T}_1 + \tilde{T}_2 + \ldots) |\Psi\rangle$, where the $\tilde{T}_i$ represent the operators that form all the $i$-th substitutions in equation 1.46. The method is denoted CCD if only double excitations are included, CCSD if singles and doubles are included. The method is size consistent, but the variational bound does not apply as the coefficients in the expansion are determined by requiring that the overlap between $(\hat{H} - E) |\Psi_{CC}\rangle$ and $|\Psi\rangle$, $|\Psi_r^a\rangle$, $|\Psi_r^s^a^b\rangle$, $\ldots$ be zero.
Scaling

CISD and CCSD scale as $\sim N^6$. 

### 1.5 Density Functional Theory

Density Functional Theory (DFT) approaches the $N$-electron problem using the electron density as the basic variable instead of the full $N$-electron wavefunction. This is justified by the Hohenberg-Kohn theorem as follows.

The electron density $\rho(r)$, is defined as

$$
\rho(r) = N \int dr_2...dr_N \Psi^*(r,r_2,...,r_N)\Psi(r,r_2,...,r_N)
$$

where all the spin variables have been integrated over.

The ground state density $\rho$ fixes number of electrons in the system since

$$
\int dr \, \rho(r) = N
$$

The potential between the nuclei and electrons, also called the external potential $v(r) = V_{Ne}(r)$ is determined by $\rho(r)$ apart from an additive constant. The proof is by contradiction. Consider two systems, one with with a Hamiltonian containing the external potential $v(r)$ and having a ground state $H\ket{\Psi} = E\ket{\Psi}$ and the other with identical Hamiltonian except for a different external potential $v'(r)$ and ground state $H'\ket{\Psi'} = E'\ket{\Psi'}$. For simplicity the ground states are taken as non-degenerate. If the primed system had an identical ground state density as the unprimed system, then

$$
E < \langle \Psi' | H | \Psi' \rangle = \langle \Psi' | H' | \Psi' \rangle + \langle \Psi' | (H - H') | \Psi' \rangle = E' + \int dr \rho(r) [v(r) - v'(r)]
$$

(1.49)
$E' < \langle \Psi | H' | \Psi \rangle = \langle \Psi | H | \Psi \rangle + \langle \Psi' | (H' - H) | \Psi' \rangle = E - \int d\mathbf{r} \rho(\mathbf{r}) [v(\mathbf{r}) - v'(\mathbf{r})]$ (1.50)

where the variational theorem has been applied to $\langle \Psi' | H | \Psi' \rangle$ and $\langle \Psi | H' | \Psi \rangle$. Adding equations 1.49 and 1.50 then implies that $E_0 + E'_0 < E'_0 + E_0$, a contradiction. So $v(\mathbf{r})$ is uniquely determined by $\rho(\mathbf{r})$, up to a constant. Since $v(\mathbf{r})$ and $N$ fix the Hamiltonian, $\rho(\mathbf{r})$ fixes the $N$-electron problem.

**Variational Theorem**

The energy of a system can be written as a functional of the density

$$E[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) v(\mathbf{r}) + F[\rho]$$ (1.51)

where $F[\rho]$ represents the kinetic and Coulomb contributions to the total energy. It is an *universal* functional of $\rho$ in the sense that it is independent of the external potential, and information on the number of particles is contained in the density through equation 1.48. The exact functional form of $F[\rho]$ is not explicitly known, but in principle, once it is known it can be used for any system.

The variational theorem in its original form, in equation 1.27, states that the energy functional $E[\Psi]$ has a minimum at the ground state of the system $\Psi_0$. Implicit in that proof is that the number of particles, $N$ remains constant as the wavefunction is varied. A variational principle in terms of trial densities was presented in Hohenberg and Kohn, but was limited to densities that corresponded to some external potential, *v*-representable densities. Not all densities are *v*-representable and general conditions for *v*-representability are not known. The search space of functions in the variational theorem should include any reasonable density that corresponds to a physically sensible wavefunction.

The limitation on *v*-representability is eliminated by the constrained search formulation of the variational problem. The search space in the constrained search formalism consists of $N$-
representable densities, densities that map to a \(N\)-electron antisymmetric wavefunction. The conditions on the density for \(N\)-representability are simple and are satisfied for any reasonable density.\(^{18}\)

\[
\rho(r) \geq 0 \quad \int dr \rho(r) = N \quad \int dr |\nabla \rho(r)|^{1/2} < \infty
\] (1.52)

Every antisymmetric wavefunction corresponds to unique density, as defined through equation 1.47, but there are many different wavefunctions that may give the same density. Following Kohn\(^{19}\) and taking \(\tilde{\rho}\) as a trial density, let \(\tilde{\Psi}_\alpha^{\tilde{\rho}}\) be the set of wavefunctions with this \(\tilde{\rho}\)

\[
E_v[\tilde{\rho}] = \min_\alpha \langle \tilde{\Psi}_\alpha^{\tilde{\rho}} | \hat{H} | \tilde{\Psi}_\alpha^{\tilde{\rho}} \rangle = \int dr v(r) \tilde{\rho}(r) + F[\tilde{\rho}(r)]
\] (1.53)

where

\[
F[\rho] = \min_\alpha \langle \tilde{\Psi}_\alpha^{\rho} | \hat{T}_e + \hat{V}_{\text{ee}} | \tilde{\Psi}_\alpha^{\rho} \rangle
\] (1.54)

and the minimization is over all antisymmetric wavefunctions that map to the density \(\rho\). Next, minimize over all \(\tilde{\rho}\),

\[
E = \min_{\rho} E_v[\tilde{\rho}]
\] (1.55)

The minimum is the ground state density, or one of the ground state densities in the case of a degenerate ground state.

**The Kohn-Sham method**

The explicit functional form of \(F[\rho]\) in equation 1.51 is unknown. In the Kohn-Sham approach a non-interacting reference system is introduced. This system has exact solutions in terms of a single
Slater determinant made up of single particle orbitals. The energy $E[\rho]$ is decomposed in the form

$$ E[\rho] = T_s[\rho] + \int dr v(r) \rho(r) + \frac{1}{2} \int dr dr' \frac{\rho(r)\rho(r')}{|r-r'|} + E_{xc}[\rho] \tag{1.56} $$

where the first term is the kinetic energy of a system where there are no electron-electron interaction terms. This non-interacting $T_s[\rho]$ is not equivalent to the actual kinetic energy functional, $T[\rho]$, of the interacting system, but it is assumed they are close. The second term is the exact functional for the external potential and the third term is the classical Coulomb energy. This equation serves to define $E_{xc}[\rho]$, which is called the exchange-correlation energy. The functional form of $E_{xc}[\rho]$ is not exactly known, but it contains the non-classical part of $V_{ee}$ and the difference between the kinetic energy of the interacting system and the non-interacting system,

$$ E_{xc}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - \int dr dr' \frac{\rho(r)\rho(r')}{|r-r'|} \tag{1.57} $$

The energy functional in equation 1.56 must be minimized along with a constraint enforcing the normalization, equation 1.48,

$$ \delta E[\rho] - \mu \left[ \int dr \rho(r) - N \right] = 0 \tag{1.58} $$

which leads to

$$ \mu = \frac{\delta T_s[\rho]}{\delta \rho(r)} + v(r) + \int dr' \frac{\rho(r')}{|r-r'|} + V_{xc}(r) \tag{1.59} $$

with the definitions

$$ V_{KS}(r) = v(r) + \int dr' \frac{\rho(r')}{|r-r'|} + V_{xc}(r) \tag{1.60} $$

and

$$ V_{xc}(r) = \frac{\delta E_{xc}[\rho]}{\delta \rho(r)} \tag{1.61} $$

Equation 1.59 represents a non-interacting system of electrons moving in the external potential.
V_{KS}(r). Each of the potential terms in 1.60 is simply a function of \( r \), and the kinetic energy term for non-interacting electrons is known. Written in terms of wavefunctions, 1.59 corresponds to

\[
\left[ -\frac{1}{2} \nabla^2 + V_{KS}(r) \right] \psi_i(r) = \varepsilon_i \psi_i(r) \tag{1.62}
\]

which has an exact ground state wavefunction which is a single determinant, \( |\Psi_s\rangle = |\psi_1 \psi_2 \ldots \psi_N\rangle \).

The exchange-correlation functional is not known explicitly, however several approximations for the functional are available. A commonly used functional is B3LYP,\(^{20}\) which combines an exchange functional due to Becke\(^{21}\) and a correlation functional due to Lee, Yang and Parr.\(^{22}\)

### Scaling of KS/DFT

The computational difficulty of KS/DFT using B3LYP scales as\(^{15} \sim N^4\) with no further simplification employed.
Chapter 2

$N$-representability

2.1 Density matrices

The antisymmetry of the wavefunction under exchange of electrons allows expectation values for
the operators in equation 1.18 to be written in terms of simpler objects called density matrices.$^{23-25}$

The electronic kinetic energy term of equation 1.18 and the nuclear-electronic potential term in
equation 1.18 each reference coordinates for a single electron at a time. The expectation value of
each such operator, called a 1-body or 1-electron operator is

$$\langle \hat{O}_i^{(1)} \rangle = \int d\mathbf{x}_1 \cdots d\mathbf{x}_N \Psi^\ast (\mathbf{x}_1, \ldots, \mathbf{x}_N) \hat{O}^{(1)}(\mathbf{x}_i) \Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)$$  \hspace{1cm} (2.1)

which after relabeling $\mathbf{x}_i$ to $\mathbf{x}$, and permuting $\mathbf{x}$ to the first argument is

$$\langle \hat{O}_i^{(1)} \rangle = \int d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \Psi^\ast (\mathbf{x}, \mathbf{x}_2, \ldots, \mathbf{x}_N) \hat{O}^{(1)}(\mathbf{x}) \Psi(\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_N)$$  \hspace{1cm} (2.2)

The expectation value of each of the 1-body terms in equation 1.18 takes this identical form so
the sum over different electrons reduces to $N$ times the value of the single integral in equation 2.2.

Expectation values of operators in a $N$-electron state can be expressed by extracting the $\mathbf{x}$
dependent parts of the integral in terms of density matrices. For one-electron operators define the
\textit{one-body reduced density matrix} as
\[ \gamma(x'; x) = N \int d x_2 \cdots d x_N \Psi^*(x', x_2, \ldots, x_N) \Psi(x, x_2, \ldots, x_N) \] (2.3)

where the normalization \( N \) is included as a convenience, as there are always \( N \) terms in the sums
in the Hamiltonian. We have
\[ \int d x \gamma(x; x) = N \] (2.4)

All one-body operator expectation values can be then written as a trace over the operator acting
on \( \gamma \).
\[ \langle \sum_i \hat{O}_i^{(1)} \rangle = \int d x \left[ O(x) \gamma(x'; x) \right]_{x' = x} = \int d x' \delta^3(x' - x) \int d x O(x) \gamma(x'; x) \] (2.5)

For an operator that acts on two electrons symmetrically, such as the electron-electron repulsion
term in equation 1.18
\[ \langle \hat{O}_{ij}^{(2)} \rangle = \int d x_1 d x_2 \cdots d x_N \Psi^*(x_1, \cdots, x_N) O(x_i, x_j) \Psi(x_1, \cdots, x_N) \] (2.6)
a density matrix
\[ \Gamma(x_1', x_2'; x_1, x_2) = \binom{N}{2} \int d x_3 \cdots d x_N \Psi^*(x_1', x_2', x_3, \ldots, x_N) \Psi(x_1, x_2, x_3 \ldots x_N) \] (2.7)
can be defined in a similar way so that the expectation value can be written
\[ \langle \sum_{i<j} \hat{O}_{ij}^{(2)} \rangle = \int d x_1 d x_2 \left[ \hat{O}^{(2)}(x_1, x_2) \Gamma(x_1', x_2'; x_1, x_2) \right]_{x_1' = x_1, x_2' = x_2} \]
\[ = \int d x_1' d x_2' \delta^3(x_2' - x_2) \delta^3(x_1' - x_1) \int d x_1 d x_2 \hat{O}^{(2)}(x_1, x_2) \Gamma(x_1', x_2'; x_1, x_1) \] (2.8)

where the normalization of \( \Gamma(x_1', x_2'; x_1, x_2) \) has been chosen appropriately for an operator sym-
metric in two electron coordinates.

In a similar way, density matrices can be defined for operators acting on any number of variables. For a symmetric $p$-electron operator

$$
\Gamma(p)(x_1', x_2', \ldots, x_p'; x_1, x_2, \ldots, x_p) = \binom{N}{p} \int dx_{p+1} \cdots dx_N \Psi^*(x_1', x_2', \ldots, x_p', x_{p+1}, \ldots, x_N) \Psi(x_1, x_2, \ldots, x_p, x_{p+1}, \ldots, x_N) \tag{2.9}
$$

The diagonal elements of the density matrices have a simple physical interpretation

$$
\gamma(x_1; x_1) dx_1 = N \cdot P_1 \tag{2.12}
$$

where $P_1$ is the probability of finding one particle with spin $\sigma_1$ in volume $d\mathbf{r}$ around $\mathbf{r}$ and

$$
\Gamma(x_1, x_2; x_1, x_2) dx_1 dx_2 = N_{\text{pairs}} \cdot P_{12} \tag{2.13}
$$

where $P_{12}$ is the probability of finding one particle with spin $\sigma_1$ in volume $d\mathbf{r}_1$ at $\mathbf{r}_1$ and another particle with spin $\sigma_2$ in volume $d\mathbf{r}_2$ at $\mathbf{r}_2$.

Note that the diagonal elements vanish if two or more indices are equal, $\Gamma(x_1, x_1; x_1, x_1) = 0$ leaving what is called a Fermi hole as is required by the Pauli principle.$^{26,27}$

Density matrices are a characteristic of the system. They are independent of any particular observable. Once the density matrices are known they allow calculation of expectation values of
CHAPTER 2. N-REPRESENTABILITY

any operator acting on the corresponding number of variables.

The expectation value of the electronic Hamiltonian written in terms of density matrices is

\[ \langle \hat{H} \rangle = \int d\mathbf{x} [\nabla^2 \gamma(x'; \mathbf{x})]_{x' = x} \]

\[ + \sum_A \int d\mathbf{x} \gamma(x; \mathbf{x}) \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \]

\[ + \int d\mathbf{x}_1 d\mathbf{x}_2 \Gamma(x_1, x_2; x_1, x_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \]  

(2.14)

where only \( \Gamma(x'_1, x'_2; x_1, x_2) \) is necessary as \( \gamma(x'; \mathbf{x}) \) is obtainable from \( \Gamma(x'_1, x'_2; x_1, x_2) \).

The density matrices are much simpler objects than the full N-electron wavefunction. The
N-electron wavefunction depends on N electron coordinates while the 1-body and 2-body density
matrices depend only on two and four electron coordinates, respectively.

In principle varying equation 2.14 with respect to the \( \Gamma \) matrix would deliver the ground
state energy and ground state density matrices. Note however that the space of test functions
\( \tilde{\Gamma}(x'_1, x'_2; x_1, x_2) \) in the variational equation must be constrained so that all the \( \tilde{\Gamma} \) are derivable
from an appropriately antisymmetric N-electron wavefunction. If the space is not so constrained
solutions may not be physical. Density matrices that are derivable from a N-electron wavefunction
which is quantum mechanically valid are called N-representable.

In the case of the Hartree-Fock method the first order density matrix \( \gamma(x'; \mathbf{x}) \) determines all the
higher-order matrices.\(^25\)

2.2 Necessary and sufficient conditions for N-representability

The wavefunction can be expanded in a complete, discrete set of orthonormal functions,

\[ \psi(\mathbf{x}) = \sum_k c_k \psi_k(\mathbf{x}) \]  

(2.15)
then

\[ \Psi(x_1, \ldots, x_N) = \sum_{k_1, \ldots, k_N} c_{k_1, \ldots, k_N} \psi_{k_1}(x_1) \ldots \psi_{k_N}(x_N) \]  

(2.16)

Since \( \Psi \) is antisymmetric on all its coordinates \( c_{k_1, \ldots, k_N} \) is antisymmetric on all its indices and satisfies the normalization condition

\[ \sum_{k_1, \ldots, k_N} |c_{k_1, \ldots, k_N}|^2 = 1 \]  

(2.17)

Define \( c_K = (N!)^{1/2} c_{k_1, \ldots, k_N} \) for ordered configurations, \( k_1 < k_2 < \ldots < k_N \). The expansion can be rewritten as

\[ \Psi(x_1, \ldots, x_N) = \sum_K c_K \Psi_K(x_1, \ldots, x_N) \]  

(2.18)

where \( \Psi_K(x_1, \ldots, x_N) = (N!)^{-1/2} \det[\psi_{k_1}, \ldots, \psi_{k_N}] \). In other words, any antisymmetric wavefunction can be expanded in terms of a series of Slater determinants over all ordered configurations \( K \).

The normalization condition becomes

\[ \sum_K |c_K|^2 = 1 \]  

(2.19)

The density matrices can be expanded in a similar way

\[ \gamma(x'; x) = \sum_{kl} \psi^*_k(x') \psi_l(x) \gamma_{k} \]  

(2.20)

The diagonal element \( \gamma_{kk} \) is called the charge order of orbital \( k \) and \( \gamma_{kl} \) for \( k \neq l \) is called the bond-order of the two orbitals \( k, l \). The matrix \( \gamma \) is Hermitian, since \( \gamma(x; x') = \gamma^*(x'; x) \) by 2.3, so it can be transformed to diagonal form in terms of a new basis, the basis of natural spin-orbitals.
The expansion of equation 2.20 in terms of natural spin-orbitals is

\[ \gamma(x'; x) = \sum_k n_k \chi_k^*(x') \chi_l(x) \tag{2.21} \]

Using 2.4, we have

\[ \sum_k n_k = N \tag{2.22} \]

To determine the coefficients, consider \( \gamma \) written in terms of the determinant expansion,

\[ \gamma(x'; x) = N \sum_{KL} c_{KL}^* \int dx_2 \ldots dx_N \chi_K^*(x_1, x_2, \ldots, x_N) \chi_L(x_1, x_2, \ldots, x_n) \tag{2.23} \]

where the integration is over every coordinate but the first in each determinant. This leads to the result

\[ n_k = \sum_{K} |c_K|^2 \tag{2.24} \]

where the sum is over all ordered configurations which contain orbital \( k \). Along with the normalization equation 2.19, this implies \( 0 \leq n_k \leq 1 \). Any general 1-body density matrix derived from a proper \( N \)-electron antisymmetric wavefunction then has the properties

\[ \sum_k n_k = N \quad 0 \leq n_k \leq 1 \tag{2.25} \]

\[ \gamma(x'; x) = \sum_k n_k \chi_k^*(x') \chi_k(x) \tag{2.26} \]

Any density matrix which is derived from a normalized, antisymmetric \( N \)-electron wavefunction will have these properties. These are necessary conditions for \( \gamma \) to be \( N \)-representable. Note that the expansion is over an infinite number of determinants so conditions 2.25 are referred to as the ensemble \( N \)-representability conditions.

Coleman\textsuperscript{28} proved that these conditions are also sufficient, that is, an arbitrary function satis-
fying 2.25 and 2.26 can be obtained from a normalized, antisymmetric \(N\)-electron wavefunction.

A general one body function of the form \(f(x'; x)\) will not necessarily map back to a quantum mechanically proper \(N\)-electron antisymmetric wavefunction. Only functions that satisfy equations 2.25 and 2.26, the \(N\)-representability conditions, are derivable from a proper \(N\)-electron wavefunction. This correspondence is illustrated in figure 2.1.

![Figure 2.1](image_url)

**Figure 2.1:** All \(N\)-representable density matrices \(\gamma\) map to quantum mechanically valid antisymmetric \(N\)-electron wavefunctions \(\Psi\), and all \(N\)-electron wavefunctions map to a \(N\)-representable density matrix. An arbitrary function \(f\) will not generally map to a valid wavefunction. Adapted from Polkosnik and Massa.\(^{29}\)

If the expansion is limited to a single determinant then \(N\) of the \(n_k\) are exactly 1 with the remainder being zero. The conditions become

\[
\begin{align*}
\text{tr} \gamma &= N \\
\gamma^2 &= \gamma 
\end{align*}
\]  

(2.27)

\[
\gamma(x'; x) = \sum_{k=1}^{N} \chi_k^*(x) \chi_k(x') 
\]  

(2.28)

Löwdin showed that these conditions are sufficient for single-determinant \(N\)-representability.

Sufficient conditions on the 2-body matrix are known but are more complicated.\(^ {30}\) In this thesis the focus is on single-determinant \(N\)-representability.

For a restricted Hartree-Fock calculation, there are a finite set of molecular orbitals \(\phi\) which are
expanded in a finite basis $\psi$ where the orbitals used in the expansion are not necessarily orthonormal. In a restricted calculation the basis orbitals are additionally taken to be identical for spin up and spin down electrons. The $N/2$ lowest energy orbitals are taken to be the only occupied orbitals, $n_1, \ldots n_{N/2} = 1$, $n_{k>N/2} = 0$ and the wavefunction consists of a single determinant made up of these orbitals. Taking these considerations into account, the restricted one-body density matrix is a normalized projector

$$\int dr \, \rho(r, r) = N$$  \hspace{1cm} (2.29)

$$\rho(r', r) = \int dr'' \, \rho(r', r'') \rho(r'', r)$$  \hspace{1cm} (2.30)

Writing $\phi = C \psi$,

$$\rho(r', r) = 2 \text{tr} \phi(r) \phi^\dagger(r') = 2 \text{tr} C \psi(r) \psi^\dagger(r') C^\dagger = 2 \text{tr} R \psi(r) \psi^\dagger(r')$$  \hspace{1cm} (2.31)

The matrix $R = C^\dagger C$ is the density matrix expressed in the non-orthonormal $\psi$ basis.

The density matrix $P$ is defined in terms of the $R$ density matrix through the square root of the overlap matrix $S$. The square root of $S$ is used to obtain a matrix which is expressed in an orthonormal basis.

$$S = \int dr \, \psi(r) \psi^\dagger(r)$$  \hspace{1cm} (2.32)

$$P = S^{1/2} \, R \, S^{1/2}$$  \hspace{1cm} (2.33)

To enforce single-determinant $N$-representability the following conditions on $P$ are sufficient

$$2 \text{tr} P = N \quad P^2 = P$$  \hspace{1cm} (2.34)
Chapter 3

Clinton Equations

3.1 The Clinton equations for general constraints

In the KEM method, KEM density matrices for the full molecule will be built from the density matrices of single and double kernels. These initial KEM density matrices will not in general satisfy $N$-representability. An iterative procedure due to Clinton purifies matrices that are not projectors, producing projector matrices that also satisfy other constraints. Examples of constraints on the density matrix are constraints to fit expectation values from observations, such as from a X-ray crystallography experiment,\textsuperscript{31,32} or purely mathematical, such as the normalization of the density matrix. The resulting normalized projector matrices are single-determinant $N$-representable.\textsuperscript{33}

A projector matrix is idempotent, $P^2 = P$. In order to enforce this property on a given matrix McWeeny\textsuperscript{34} proposed minimizing the scalar expression $\text{tr}((P^2 - P)^2)$. This is the square of the Frobenius norm of the difference between $P^2$ and $P$, $\|P^2 - P\|^2$. Varying this expression with respect to $P$,

$$\delta \text{tr}(P^2 - P)^2 = 2 \text{tr}(2P^3 - 3P^2 + P) \delta P = 0 \quad (3.1)$$
which leads to an iterative equation for purifying a matrix which is nearly a projector of the form

\[ P_{n+1} = 3P_n^2 - 2P_n^3 \]  

(3.2)

To generalize this expression to include constraints based on one-electron operator expectation values, which are of the form \( \langle \hat{O} \rangle = \text{tr} \hat{P} \hat{O} \) the method of Lagrange multipliers is used. Introducing Lagrange multipliers \(-2\lambda_k\) to equation 3.1 for each of \(k\) constraints, the quantity to be minimized becomes

\[
\text{tr} (P^2 - P)^2 - \sum_k 2\lambda_k \text{tr} \hat{P} \hat{O}
\]

(3.3)

which leads to

\[
2P^3 - 3P^2 + P - \sum_k \lambda_k O_k
\]

(3.4)

This equation can be solved using the a fixed point iterative procedure. Starting with a matrix \(P_0\), the equation

\[
P_{n+1} = 3P_n^2 - 2P_n^3 + \sum_k \lambda_k^{(n)} O_k
\]

(3.5)

is iterated until \(\text{tr} (P^2 - P)^2\) is below some tolerance.

The Lagrange multipliers at each step are obtained by satisfying the constraint equations \(\text{tr} P_{n+1} O_k = \langle \hat{O}_k \rangle\). The Lagrange multipliers approach zero as the matrix \(P\) approaches idempotency.

\[
\begin{pmatrix}
\text{tr} \hat{O}_1 \hat{O}_1 & \text{tr} \hat{O}_1 \hat{O}_2 & \cdots & \text{tr} \hat{O}_1 \hat{O}_k \\
\vdots & \vdots & \cdots & \vdots \\
\text{tr} \hat{O}_k \hat{O}_1 & \text{tr} \hat{O}_k \hat{O}_2 & \cdots & \text{tr} \hat{O}_k \hat{O}_k \\
\end{pmatrix}
\begin{pmatrix}
\lambda_1^{(n)} \\
\vdots \\
\lambda_k^{(n)} \\
\end{pmatrix}
= 
\begin{pmatrix}
\langle \hat{O}_1 \rangle - \text{tr}[(3P_n^2 - 2P_n^3) \hat{O}_1] \\
\vdots \\
\langle \hat{O}_k \rangle - \text{tr}[(3P_n^2 - 2P_n^3) \hat{O}_k] \\
\end{pmatrix}
\]

(3.6)

The resulting projector determines a set of orbitals which define a single Slater determinant, which guarantees that the corresponding energy satisfies the variational principle.
3.2 Single-determinant $N$-representability via the Clinton equations

In KEM we are interested in a single constraint, a constraint that enforces correct normalization of the density matrix. For restricted density matrices the normalization condition is $2 \text{tr} \, P = N$. Extremizing the squared norm of the difference between $P^2$ and $P$ subject to a normalization constraint,

$$
\delta \left[ \text{tr} \left( (P^2 - P)^2 \right) - 2 \lambda (\text{tr} P - N/2) \right] = 0
$$

leads to an iterative equation providing a normalized projector

$$
P_{n+1} = 3P_n^2 - 2P_n^3 + \lambda_n \mathbb{1}
$$

with the Lagrange multiplier $\lambda$ at each step of the iteration given by

$$
\lambda_n = \left[ N/2 - \text{tr}(3P_n^2 + 2P_n^3) \right] / \text{tr} \mathbb{1}
$$

The basic KEM energy expression does not satisfy the variational bound, as the variational theorem only holds for energies determined from trial wavefunctions and the kernel expansion for energy does not involve wavefunctions for the $N$-electron system. However the variational bound can be recovered in the kernel expansion by using the KEM expansion for the density and applying the Clinton equations. By starting with an initial density matrix given by the generalization of the KEM expansion for the density matrix, as defined in equation 5.12, equations 3.8 and 3.9 can be used to convert an initial $P_0$ to a normalized projector. This projector corresponds to a single Slater determinant, the energy of which satisfies the variational principle.
Chapter 4

X-ray Crystallography

A crystal is a periodic arrangement of identical structural units throughout space. The structural units can consist of single or multiple atoms. This arrangement can be described by attaching a fixed basis of atoms to a periodic arrangement of points in space, a lattice. \(^{35}\) The periodicity of the lattice is expressed mathematically in terms of a crystal translation vector

\[
t = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3
\]  

where the \(n_i\) are integers and the \(\mathbf{a}_i\) are the primitive axes. The primitive axes define a parallelepiped called the primitive cell.

Lattices are characterized by their symmetry. In addition to the lattice translation symmetry there may be rotation and reflection symmetries that carry the crystal into itself (point operations), or combined lattice and translation symmetries. The possible lattice rotation symmetries include rotation by \(2\pi, 2\pi/2, 2\pi/3, 2\pi/4\) and \(2\pi/6\). The point symmetries also include mirror reflections through a plane and inversions, \(\mathbf{r} \rightarrow -\mathbf{r}\).

The basis is specified by vectors giving the positions of atoms relative to a fixed origin, usually
taken to be a lattice point, in terms of the primitive axes,

\[ r_i = x_i a_1 + y_i a_2 + z_i a_3 \]  \hspace{1cm} (4.2)

with \(0 \leq x_i \leq 1, \ 0 \leq y_i \leq 1, \ 0 \leq z_i \leq 1.\)

### 4.1 X-ray diffraction

If radiation which is incident on a crystal has a wavelength which is of the order of or less than the length of the primitive lattice vectors, \(\lambda \lesssim a\), there is a possibility of diffraction. Constructive interference from the scattered waves is possible leading to diffracted beams in various directions. These beams will have varying intensities which depend on the structure of the basis. Typical crystal lattice vectors have \(a \sim 1\,\text{Å} – 10\,\text{Å}\) which is in the X-ray region of the electromagnetic spectrum.

The crystal lattice is periodic with translation vector 4.1, so any physical properties will be invariant under \(t\) translations as well.

The electron density determines the scattering, and will be periodic, \(\rho(r + t) = \rho(r)\). The density can be expanded in Fourier series,

\[ \rho(r) = \sum_q \rho_q e^{iq \cdot r} \]  \hspace{1cm} (4.3)

The vector \(q\) is a reciprocal lattice vector,

\[ q = m_1 b_1 + m_2 b_2 + m_3 b_3 \]  \hspace{1cm} (4.4)

where

\[ b_1 = 2\pi \frac{a_2 \times a_3}{a_1 \cdot a_2 \times a_3} \quad b_2 = 2\pi \frac{a_3 \times a_1}{a_1 \cdot a_2 \times a_3} \quad b_3 = 2\pi \frac{a_1 \times a_3}{a_1 \cdot a_2 \times a_3} \]  \hspace{1cm} (4.5)
The phase factor for rays that are elastically scattered from two elements of the local electron concentration separated by a vector \( \mathbf{r} \) is \( e^{i\Delta \mathbf{k} \cdot \mathbf{r}} \) with \( \Delta \mathbf{k} = \mathbf{k'} - \mathbf{k} \) as illustrated in figure 4.1. The amplitude of the waves depends on the local electron concentration. The scattering amplitude is

\[
F = \int d\mathbf{r} \, \rho(\mathbf{r}) e^{-i\Delta \mathbf{k} \cdot \mathbf{r}} = \sum_q \int d\mathbf{r} \, \rho_q e^{iq(\mathbf{q} - \Delta \mathbf{k}) \cdot \mathbf{r}} \quad (4.6)
\]

The argument of the exponential is zero for \( \mathbf{q} = \Delta \mathbf{k} \) and \( |\mathbf{k'}| = |\mathbf{k}| \) for elastic scattering. This gives a condition for diffraction

\[
2\mathbf{k} \cdot \mathbf{q} = \mathbf{q}^2 \quad (4.7)
\]

so the set of reciprocal lattice vectors determines the locations of the diffraction peaks.

When the condition is satisfied the scattering amplitude is

\[
F_q = \int d\mathbf{r} \, \rho(\mathbf{r}) e^{iq \cdot \mathbf{r}} \quad (4.8)
\]
4.2 Structure determination

Given scattering data $F_{q}^{\text{expt}}$, the position of the atoms in a crystal are commonly determined by fitting the observed scattering data to $F_{q}^{\text{calc}}$ which is calculated based on some model of the electron density.

One possibility is to assume the electron density to be a sum of spherical atomic electron densities.

$$\rho(r) = \sum_{i} \rho_{i}^{s}(r)$$  \hspace{1cm} (4.9)

The centers of the spheres are adjusted to minimize the $R$-factor,

$$\delta R = \delta \left( \sum_{q} w_{q} \left[ F_{q} - F_{q}^{\text{expt}} \right]^{2} \right) = 0$$  \hspace{1cm} (4.10)

where the $w_{q}$ are weighting factors that take the experimental errors into account. Data at a $q$ with higher experimental uncertainty is weighed less in the sum.

Chemical bonding introduces non-spherical densities. These can be modeled by including a multipole expansion of the density. A commonly used model is due to Hansen and Coppens,\textsuperscript{36}

$$\rho_{\text{atomic}}(r) = P_{c}\rho_{\text{core}}(r) + P_{v}\kappa^{3}\rho_{\text{valence}}(\kappa^{\prime}r) + \sum_{l=0}^{4} \kappa^{l}R_{l}(\kappa^{\prime}r) \sum_{m=-l}^{l} P_{lm}y_{lm}(r/r)$$  \hspace{1cm} (4.11)

where $y_{lm}$ are real spherical harmonics and $P_{c}, P_{v}, P_{lm}$ are population coefficients.

4.3 Quantum Crystallography

Quantum crystallography is a multidisciplinary field that works at the interface of quantum chemistry and crystallography, using information from one to inform the other.\textsuperscript{37} From one direction, data from X-ray scattering can be useful to quantum chemical calculations. For example, X-ray scattering contains information on the electron density including electron correlation. From the
other direction, quantum chemical calculations can give insight into crystallographic techniques, such as refining the density models used in structure determination, as in equation 4.11.

A framework for quantum crystallography was developed by Clinton and collaborators.\textsuperscript{33} The framework is based on using the Clinton equations with constraints which are set by the experimental data. The constraints are

$$\text{tr} Pf(q) = F_q^{\text{expt}}$$

(4.12)

where $f(q)$ is the matrix representative of $e^{iq \cdot r}$ in the chosen basis.

Based on this framework Clinton and Massa demonstrated the extraction of a density matrix from simulated X-ray diffraction data in 1972.\textsuperscript{31} The first extraction of a density matrix from experimental scattering data was by Massa et. al. for metallic beryllium.\textsuperscript{32}

In all the above studies single-determinant $N$-representability was enforced by the Clinton equations along with experimental constraints of the form of equation 4.12. This is required to obtain quantum mechanically valid results.

**KEM, $N$-representability and Quantum Crystallography**

For large molecules the technique used above becomes under-determined. The number of matrix elements in $P$ rises as the square of the number of atoms while the X-ray data increases directly with the number of atoms. However, in principle, given the atomic positions obtained from the standard structure determination procedure, KEM can deliver the full quantum mechanics of large molecules with good accuracy.\textsuperscript{38} The density, 1-body density matrix, and 2-body density matrix are given by the KEM expansions

$$\rho = \sum_{i<j}^K \rho_{ij} - (K - 2) \sum_{i=1}^K \rho_i$$

(4.13)
\[ \rho_1 = \sum_{i<j}^K \rho_{1ij} - (K - 2) \sum_{i=1}^K \rho_{1i} \]  
(4.14)

\[ \rho_2 = \sum_{i<j}^K \rho_{2ij} - (K - 2) \sum_{i=1}^K \rho_{2i} \]  
(4.15)

The quantum mechanics of large molecules such as proteins is still not routinely accessible by quantum mechanical methods. Given atomic positions from a X-ray crystallography experiment, KEM can deliver the full set of density matrices which are required for calculation of expectation values and thus the full quantum mechanics of crystallized molecules of any size.
Chapter 5

The Kernel Energy Method

5.1 The KEM equations

The strategy of the kernel energy method is to decompose a molecular system into subsets and unions of these subsets for the purposes of computation. An individual subset is called a kernel or a single kernel. The single kernels must all be disjoint. The union of two kernels is called a double kernel, the union of three a triple kernel and so on. The union of all the single kernels recovers the full molecular system. Whenever there is a bond spanning two kernels, the atoms participating in the bond in both kernels are capped with a hydrogen atom in order to preserve the nature of the bond. In general kernels are chosen so that only single bonds are cut. Otherwise the choice of the particular kernel decomposition used in a calculation is left up to the judgment of the researcher using the method.

The results of calculations on the collection of single, double and higher order kernels is combined in a way that is particular to KEM to obtain results pertaining to the full system. The double kernels and higher order kernels capture the interactions between atoms in the corresponding kernels. Any chemical model can be used for KEM kernel calculations.

The kernels consist of fewer atoms than the entire molecule and so the kernel energy method
scales slower than *ab-initio* methods on the full molecule. Each kernel and double kernel is treated completely independently for the purposes of computation so the method is very well suited to parallel computation and is efficient in memory space. The accuracy increases as the size of the kernels increases.\(^{40}\)

The fundamental idea behind the KEM expansion for the energy is that full molecule energy is the sum of the energies of all single kernels together with the interaction energies between pairs of kernels. This double kernel sum captures the energy of interaction between all the single kernels. In summing the results of the double kernels the contributions of single kernels are overcounted and the overcounting is subtracted from double kernel sum. The KEM energy to second order is defined as

\[E_{\text{KEM}} = K \sum_{i < j} E_{ij} - (K - 2) \sum_i E_i\]  \hspace{1cm} (5.1)

In this equation \(E_{\text{KEM}}\) is the KEM approximation to the full molecule energy. The first term on the right side of the equation is the sum of the energies of all the double kernels, with each double kernel contributing \(E_{ij}\). The \(E_i\) represent energies of single kernels. Each of the single kernels is over counted \((K - 2)\) times in the double kernel sum, and so the second term subtracts this overcounting. Caps are included in the double and single kernel calculations if capping of bonds was necessary. The contribution of the capping hydrogens approximately cancel out in equation 5.1. The energy sum of equation 5.1 can also be understood as a sum of all single kernel energies with all double kernel interaction energies. Defining the interaction energy between two kernels as \(I_{ij} = E_{ij} - E_i - E_j\), the sum is equivalent to

\[E_{\text{KEM}} = \sum_i E_i + \sum_{i < j} I_{ij}\] \hspace{1cm} (5.2)

This idea is characteristic of the approximation taken by KEM.

For molecules which is made up of a chain of units such as a protein a further approximation
is to only consider the chemically bonded kernels

\[ E_{\text{KEM}} = \sum_{i=1, j=i+1}^{K-1} E_{ij} - \sum_{i=2}^{K-1} E_i \]  

(5.3)

Such a KEM expansion can scale linearly with the number of kernels.\(^\text{40}\)

The KEM expansion has been generalized to include higher order kernel interactions.\(^\text{15,41}\)

KEM has been demonstrated to deliver accurate energy results for biological molecules such as DNA, RNA and proteins\(^\text{39,40,42,43}\) using the kernel expansion 5.1 and 5.3. Additionally, KEM has been shown to give good accuracy over a range of basis sets and chemical methods.\(^\text{44}\) It gives accurate densities near bond and ring critical points and the complete localization-delocalization matrix of a graphene nanoribbon\(^\text{45–47}\) and response properties of a graphene flake.\(^\text{48}\)

### 5.2 KEM and \(N\)-representability

KEM has been demonstrated to provide accurate energies. However, as a method that does not refer to a wavefunction for the entire system, the variational theorem does not apply to KEM energies. The variational bound is useful because it provides a definite limit to the energy of a system. Any energy for a wavefunction based approximation to solutions to the electronic Hamiltonian must be higher than the exact ground state energy. More importantly, there are other observables besides the energy that are of interest for molecular systems. Extending KEM to provide \(N\)-representable density matrices allows calculation of these properties. Moreover, once a \(N\)-representable density matrix is found, expectation values of all operators are easily calculated. All one body operators have expectation values given by simply evaluating the trace of the product of the matrix representative of the operator and the density matrix, \(\langle \hat{O} \rangle = \text{tr} \hat{P} \hat{O} \).

In order to adapt expression 5.1 to provide a full molecule density matrix, the density matrices of each single and double kernel must be adapted to the full molecule basis. Each single and double
kernel density matrix refers to a subset of the full molecule basis. Using lowercase symbols to refer to the single and double kernel density matrices in these subsets of the full molecule matrix space, the single kernel density matrix $r_i$ has basis indices that only include basis functions corresponding to the atoms in that kernel. The double kernel density matrix has indices that only refer to basis functions corresponding to atoms in that double kernel, $r_{ij}$. To adapt the matrices $r_i$ and $r_{ij}$ to the full molecule basis, the elements from each matrix are placed at appropriate places in a matrix the size of the full molecule density matrix with all other elements set to zero. These are called the augmented single and double kernel matrices and will be denoted by uppercase symbols $R_i$ and $R_{ij}$.

The augmented form of $r_i$ is

$$
R_i = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & r_i & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\end{pmatrix}
$$

In particular this would be the form for $R_2$ in a kernel expansion involving four kernels. The entire single kernel density matrix $r_i$ is placed as a block into the appropriate position in a matrix which is the size of the full molecule matrix.

Written explicitly in terms of basis indices the augmented matrix of a single kernel is

$$
(R_i)_{\mu+i,\nu+i} = \begin{cases}
(r_i)_{\mu \nu}, & \mu = 1 \ldots n_i, \nu = 1 \ldots n_i \\
0, & \text{otherwise}
\end{cases}
$$
The augmented form of matrix $r_{ij}$ is

$$R_{ij} = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & (r_{ij})^{ii} & 0 & 0 & (r_{ij})^{ij} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & (r_{ij})^{ji} & 0 & 0 & (r_{ij})^{jj} & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \tag{5.6}$$

This would be the form for $R_{25}$ in a six kernel decomposition. The form of the augmented matrix in the case of double kernels is more complicated than for single kernels as four different blocks of the double kernel density matrix must be placed the four different blocks corresponding to the basis set indices for the two kernels making up the double kernel.

In equation 5.6 subscripts denote kernels while superscripts denote blocks of the density matrix for that kernel. The matrix $(r_{ij})^{ii}$ is the block from the double kernel density matrix that corresponds to basis function indices for atoms only in kernel $i$ and $(r_{ij})^{jj}$ is the block referring to basis function indices for atoms only in kernel $j$. The off-diagonal blocks $(r_{ij})^{ij}$ and $(r_{ij})^{ji}$ are elements of the double kernel density matrix which refer to basis function indices from atoms in kernels $i$ and $j$. The remaining elements of the matrix are set to zero.\(^{29}\)

Written explicitly in terms of indices

$$\begin{pmatrix} R_{ij} \end{pmatrix}_{\mu+\mu, \nu+\nu} = \begin{cases} (r_{ij})_{\mu \nu}, & \mu = 1 \ldots n_i, \nu = 1 \ldots n_i \\ 0, & \text{otherwise} \end{cases} \tag{5.7}$$

$$\begin{pmatrix} R_{ij} \end{pmatrix}_{\mu+\mu, \nu+\nu} = \begin{cases} (r_{ij})_{\mu \nu}, & \mu = 1 \ldots n_j, \nu = 1 \ldots n_j \\ 0, & \text{otherwise} \end{cases} \tag{5.8}$$
\[(R_{ij})_{\mu+\mu',\nu+\nu'} = \begin{cases} (r_{ij})_{\mu\nu}, & \mu = 1 \ldots n_i, \nu = 1 \ldots n_j \\ 0, & \text{otherwise} \end{cases} \quad (5.9)\]

\[(R_{ij})_{\mu+\mu',\nu+\nu'} = \begin{cases} (r_{ij})_{\mu\nu}, & \mu = 1 \ldots n_j, \nu = 1 \ldots n_i \\ 0, & \text{otherwise} \end{cases} \quad (5.10)\]

where

\[\mu_r = \sum_{l=1}^{r-1} n_i, \quad \nu_s = \sum_{l=1}^{s-1} n_j\]

with \(n_l\) defined as the basis set size of kernel \(l\).

The overlap matrix \(S\) in the full molecule basis is obtained from the double and single kernel overlap matrices through a similar block decomposition.

Using the augmented kernel matrices, the KEM version of the full molecule density matrix is defined as

\[R_{\text{KEM}} = \sum_{i<j}^K R_{ij} - (K-2) \sum_i^K R_i\]

With this definition of \(R_{\text{KEM}}\), the matrix \(P_{\text{KEM}}\) is defined by equation 2.33. This initial \(P_{\text{KEM}}\) is normalized by construction but is not necessarily a projector. The Clinton equations in the form 3.8 and 3.9 are applied to this initial matrix, \(P_0 = P_{\text{KEM}}\) to obtain a normalized projector which is single-determinant \(N\)-representable.
Chapter 6

KEM Hartree-Fock calculations on Water Clusters

Water is ubiquitous in chemistry. The characteristics of water clusters are important in the study of cloud and ice formation and in solution chemistry. They play an essential role in the structure of biomolecules and in biochemical processes.\(^{49}\)

KEM \(N\)-representability was studied\(^{29}\) over a set of energy minimized water clusters. The number of water molecules ranged from \(N_{\text{water}} = 3 \rightarrow 20\). The structures used were taken from a study of the structure and stability of water clusters.\(^{49}\) Most of the geometries in that study consist of a fusion of tetrameric rings or a fusion of pentameric rings stabilized by hydrogen bonds. Examples of these two types of structures are illustrated in figures 6.1 and 6.2.\(^{50}\) For \(N_{\text{water}} = 6\), a cage structure and a prism structure of comparable energies were included.

Each cluster was broken into three kernels with approximately the same number of water molecules in each by cutting across hydrogen bonds. The energy, density matrix, one-electron integrals and two-electron integrals for each kernel and for the full molecule were obtained using the restricted Hartree-Fock method in a 6-31G\((d, p)\) basis set using Gaussian.\(^{14}\) The KEM energy was calculated for each structure according to equation 5.1. The initial KEM density matrix was
obtained from the sum of augmented kernel density matrices, equation 5.12. The initial density matrix was purified to a normalized projector, \( \mathbf{P}_{\text{projector}} \), using the Clinton equations, 3.8 and 3.9. The Clinton algorithm was judged to have converged if \( \text{tr}(\mathbf{P}^2 - \mathbf{P})^2 \leq 10^{-18} \). Details for each iteration of the algorithm are given in Polkosnik and Massa\textsuperscript{29} for some representative clusters. The energy of \( \mathbf{P}_{\text{projector}} \) was calculated using

\[
E[\mathbf{R}] = V_N + \sum_{\mu\nu} (T_e)_{\mu\nu} R_{\mu\nu} + (V_{Ne})_{\mu\nu} R_{\mu\nu} + \sum_{\mu\nu\rho\sigma} (V_{ee})_{\mu\nu\rho\sigma} [R_{\mu\nu} R_{\rho\sigma} - \frac{1}{2} R_{\mu\rho} R_{\nu\sigma}] \tag{6.1}
\]

where the first term is the nuclear-nuclear potential energy defined in 1.2. The remaining terms include the matrix representatives of the operators defined in equations 1.3 and 1.2. The matrix representative of the two-electron density matrix is \( R_{\mu\nu} R_{\rho\sigma} - \frac{1}{2} R_{\mu\rho} R_{\nu\sigma} \), which can be expressed in terms of the 1-electron density matrix explicitly in the case of a single determinant theory like Hartree-Fock.

The accuracy of the straightforward KEM energy, \( E_{\text{KEM}} \), is compared to the full molecule energy results, \( E_{\text{full}} \), and the energy of the purified density matrix, \( E[\mathbf{P}_{\text{projector}}] \), in table 6.1. Violations of the variational theorem show up as negative energy differences in the \( E_{\text{KEM}} - E_{\text{full}} \) columns. There are four violations of the variational bound for \( E_{\text{KEM}} \), in the cases \( N_{\text{water}} = 10, 12, 14, 15 \). In each of those cases the variational bound is satisfied for the energy associated with the density matrix on which \( N \)-representability has been enforced, with approximately the same accuracy. In all the non-violating cases the energy of the \( N \)-representable density matrix is substantially closer than \( E_{\text{KEM}} \) to the full molecule energy.

In every case presented in table 6.1 the Clinton algorithm converges to a normalized projector in only a few steps. The details regarding the convergence of the Clinton algorithm in the four cases that violated the variational bound and for one typical non-violating case are presented in Polkosnik and Massa.\textsuperscript{29}

As a further test of the methodology each of the structures was artificially scaled by a factor
of 0.95 to 0.55 in steps of 0.05 and $E_{\text{KEM}}$, $E_{\text{full}}$ and $E[P_{\text{projector}}]$ were compared. Structures scaled by factors less than 0.55 were not possible to calculate in Gaussian due to close inter-atomic distances. Out of a total of 190 structures calculated, 80 violated the variational bound. In each of those 80 cases the energy of $E[P_{\text{projector}}]$ satisfied the bound. The $N$-representable density matrix always provided an energy that satisfied the variational bound, even in the cases for strongly clashing systems scaled down by a factor of 0.55.
Table 6.1: KEM energy and energy of $N$-representable density matrix compared to full molecule Hartree-Fock energy across water clusters.

<table>
<thead>
<tr>
<th>$N_{\text{water}}$</th>
<th>$E_{\text{KEM}}$</th>
<th>$E_{\text{P\text{projector}}}$</th>
<th>$E_{\text{full}}$</th>
<th>$E_{\text{KEM}} - E_{\text{full}}$</th>
<th>$E_{\text{P\text{projector}}} - E_{\text{full}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a.u.</td>
<td>a.u.</td>
<td>a.u.</td>
<td>kcal·mol$^{-1}$</td>
<td>kcal·mol$^{-1}$</td>
</tr>
<tr>
<td>3</td>
<td>-227.985676</td>
<td>-227.988209</td>
<td>-227.988838</td>
<td>$3.16 \times 10^{-3}$</td>
<td>$1.98 \times 10^{-4}$</td>
</tr>
<tr>
<td>4</td>
<td>-303.995529</td>
<td>-304.000345</td>
<td>-304.005881</td>
<td>$4.82 \times 10^{-3}$</td>
<td>$3.02 \times 10^{-4}$</td>
</tr>
<tr>
<td>5</td>
<td>-380.001291</td>
<td>-380.005563</td>
<td>-380.008667</td>
<td>$4.59 \times 10^{-3}$</td>
<td>$2.88 \times 10^{-4}$</td>
</tr>
<tr>
<td>6 (cage)</td>
<td>-456.001906</td>
<td>-456.007658</td>
<td>-456.008667</td>
<td>$6.76 \times 10^{-3}$</td>
<td>$4.24 \times 10^{-3}$</td>
</tr>
<tr>
<td>6 (prism)</td>
<td>-456.007568</td>
<td>-456.008844</td>
<td>-456.009632</td>
<td>$2.06 \times 10^{-3}$</td>
<td>$1.30 \times 10^{-4}$</td>
</tr>
<tr>
<td>7</td>
<td>-532.011240</td>
<td>-532.018087</td>
<td>-532.019046</td>
<td>$7.81 \times 10^{-3}$</td>
<td>$4.90 \times 10^{-4}$</td>
</tr>
<tr>
<td>8</td>
<td>-608.023625</td>
<td>-608.033600</td>
<td>-608.035409</td>
<td>$1.18 \times 10^{-2}$</td>
<td>$7.39 \times 10^{-3}$</td>
</tr>
<tr>
<td>9</td>
<td>-684.038766</td>
<td>-684.041271</td>
<td>-684.042037</td>
<td>$3.27 \times 10^{-3}$</td>
<td>$2.05 \times 10^{-4}$</td>
</tr>
<tr>
<td>10</td>
<td>-760.051696</td>
<td>-760.049963</td>
<td>-760.051042</td>
<td>$-6.54 \times 10^{-4}$</td>
<td>$0.41 \times 10^{-3}$</td>
</tr>
<tr>
<td>11</td>
<td>-836.045216</td>
<td>-836.047270</td>
<td>-836.048114</td>
<td>$2.90 \times 10^{-3}$</td>
<td>$1.82 \times 10^{-4}$</td>
</tr>
<tr>
<td>12</td>
<td>-912.077473</td>
<td>-912.070075</td>
<td>-912.071346</td>
<td>$-6.13 \times 10^{-3}$</td>
<td>$-3.84 \times 10^{-3}$</td>
</tr>
<tr>
<td>13</td>
<td>-988.063484</td>
<td>-988.066065</td>
<td>-988.067674</td>
<td>$4.19 \times 10^{-3}$</td>
<td>$2.63 \times 10^{-3}$</td>
</tr>
<tr>
<td>14</td>
<td>-1064.057735</td>
<td>-1064.055613</td>
<td>-1064.056877</td>
<td>$-8.58 \times 10^{-4}$</td>
<td>$-0.54 \times 10^{-3}$</td>
</tr>
<tr>
<td>15</td>
<td>-1140.098043</td>
<td>-1140.094095</td>
<td>-1140.096400</td>
<td>$-1.64 \times 10^{-3}$</td>
<td>$-1.03 \times 10^{-3}$</td>
</tr>
<tr>
<td>16</td>
<td>-1216.095941</td>
<td>-1216.103235</td>
<td>-1216.106180</td>
<td>$1.02 \times 10^{-2}$</td>
<td>$6.42 \times 10^{-3}$</td>
</tr>
<tr>
<td>17</td>
<td>-1292.099468</td>
<td>-1292.103639</td>
<td>-1292.106019</td>
<td>$6.55 \times 10^{-3}$</td>
<td>$4.11 \times 10^{-3}$</td>
</tr>
<tr>
<td>18</td>
<td>-1368.097112</td>
<td>-1368.104786</td>
<td>-1368.107675</td>
<td>$1.06 \times 10^{-2}$</td>
<td>$6.63 \times 10^{-3}$</td>
</tr>
<tr>
<td>19</td>
<td>-1444.125210</td>
<td>-1444.127895</td>
<td>-1444.128387</td>
<td>$3.18 \times 10^{-3}$</td>
<td>$1.99 \times 10^{-4}$</td>
</tr>
<tr>
<td>20</td>
<td>-1520.137093</td>
<td>-1520.141517</td>
<td>-1520.144046</td>
<td>$6.95 \times 10^{-3}$</td>
<td>$4.36 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
Figure 6.3: KEM energy and energy of $N$-representable density matrix compared to full molecule Hartree-Fock energy across water clusters.
Chapter 7

KEM KS/DFT calculations on Noble Gas and Water Cluster

The methodology of enforcing $N$-representability on the KEM density matrix was further tested in the context of Kohn-Sham Density Functional Theory (KS/DFT). The method was applied to a simple proof-of-concept system and a single water cluster made up of twelve water molecules. The proof-of-concept system consisted of three helium atoms on a line. The distance between adjacent atoms was varied between $0.4 \leq r \leq 2.0$ where $r = R/R_{vdW}$ is the ratio of the separation to the van der Waals radius of helium, $R_{vdW} = 1.4$ Å. At small $r$ the atoms are strongly clashing while at large $r$ the atoms are essentially independent. Each atom was taken as a kernel. Calculations were done in the STO-3G and 6-31G($d,p$) bases using Gaussian. For $r < 0.4$ calculations were not possible due to close contacts. The energy and density matrix for each kernel and the full system were extracted from Gaussian. The energy of $P_{\text{projector}}$ was calculated using Gaussian. The convergence criterion used for the Clinton algorithm was $\|P^2 - P\|^2 < 1 \times 10^{-28}$. The equations converge in $n \lesssim 10$ steps for all cases except for $r = 0.4$ and $r = 0.5$ in the 6-31G($d,p$) basis. For $r = 0.4$ the initial value of $\|P^2 - P\|^2 = 2.8$, and increases without bound as the iterations proceed. For $r = 0.5$ the Clinton algorithm converges to $\|P^2 - P\|^2 = 5.95 \times 10^{-2}$ with $E[P] = -2.75$ a.u.. It remains at
this local minimum at all remaining iterations. Details of the iterations are presented in Polkosnik and Massa.\textsuperscript{51}

A summary of the results for the proof-of-concept system is presented in table 7.3 and figure 7.1 for the STO-3G basis and in table 7.4 and figure 7.2 for the 6-31G\((d, p)\) basis.

For large \(r, r > 1\) the direct KEM energy expansion and the KEM projector energy are close to the full system energy in both bases. For the STO-3G basis the accuracy of \(E_{\text{KEM}}\) steadily decreases as \(r\) decreases while \(E[P_{\text{projector}}]\) remains very accurate at all distances. At the smallest separation \(E_{\text{KEM}}\) is \(-137\text{ kcal mol}^{-1}\), much lower than \(E_{\text{full}}\) and a very large error. The largest deviation of \(E[P_{\text{projector}}]\) from \(E_{\text{full}}\) is also at the smallest separation between atoms but is only \(1.84 \times 10^{-5}\text{ kcal mol}^{-1}\), a very small error. In the 6-31G\((d, p)\) basis both \(E_{\text{KEM}}\) and \(E[P_{\text{projector}}]\) diverge from the full system energy for small \(r\), unlike the behavior for the STO-3G basis. In this basis the projector energy is more accurate than the KEM energy in most cases, but for \(R/R_{\text{vdW}} = 1.8, 1.7, 1.4, 1.3, 0.6\) the KEM energy is more accurate than the projector energy.

The KEM KS/DFT procedure was applied to a cluster of twelve water molecules obtained from Maheshwary \textit{et al.}\textsuperscript{49} The basis used was 6-31G\((d, p)\). The results for this case are presented in tables 7.1 and 7.2. The Clinton equations converge in a few steps. The energy of the \(N\)-representable density matrix is closer to the full molecule energy than the straightforward KEM energy by a factor of nearly four.

<table>
<thead>
<tr>
<th>(E_{\text{KEM}}) a.u.</th>
<th>(E_{\text{full}}) a.u.</th>
<th>(E_{\text{KEM}} - E_{\text{full}}) a.u.</th>
<th>(E_{\text{KEM}} - E_{\text{full}}) kcal mol(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-917.2993)</td>
<td>(-917.2915)</td>
<td>(-7.81 \times 10^{-3})</td>
<td>(-4.90)</td>
</tr>
</tbody>
</table>

Table 7.1: KEM energy compared to full molecule KS/DFT energy for a cluster of twelve water molecules.
Table 7.2: Energy of $N$-representable density matrix compared to full molecule KS/DFT energy for a cluster of twelve water molecules.
<table>
<thead>
<tr>
<th>$r$</th>
<th>$E_{\text{KEM}} - E_{\text{full}}$</th>
<th>$E[\mathbf{P}<em>{\text{projector}}] - E</em>{\text{full}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a.u.</td>
<td>kcal·mol$^{-1}$</td>
</tr>
<tr>
<td>0.4</td>
<td>$-2.18 \times 10^{-1}$</td>
<td>$-1.37 \times 10^{2}$</td>
</tr>
<tr>
<td>0.5</td>
<td>$-9.13 \times 10^{-2}$</td>
<td>$-5.73 \times 10^{1}$</td>
</tr>
<tr>
<td>0.6</td>
<td>$-3.82 \times 10^{-2}$</td>
<td>$-2.40 \times 10^{1}$</td>
</tr>
<tr>
<td>0.7</td>
<td>$-1.53 \times 10^{-2}$</td>
<td>$-9.59$</td>
</tr>
<tr>
<td>0.8</td>
<td>$-5.78 \times 10^{-3}$</td>
<td>$-3.63$</td>
</tr>
<tr>
<td>0.9</td>
<td>$-2.04 \times 10^{-3}$</td>
<td>$-1.28$</td>
</tr>
<tr>
<td>1.0</td>
<td>$-6.62 \times 10^{-4}$</td>
<td>$-4.15 \times 10^{-1}$</td>
</tr>
<tr>
<td>1.1</td>
<td>$-1.99 \times 10^{-4}$</td>
<td>$-1.25 \times 10^{-1}$</td>
</tr>
<tr>
<td>1.2</td>
<td>$-4.70 \times 10^{-5}$</td>
<td>$-2.95 \times 10^{-2}$</td>
</tr>
<tr>
<td>1.3</td>
<td>$-2.44 \times 10^{-6}$</td>
<td>$-1.53 \times 10^{-3}$</td>
</tr>
<tr>
<td>1.4</td>
<td>$2.47 \times 10^{-6}$</td>
<td>$1.55 \times 10^{-3}$</td>
</tr>
<tr>
<td>1.5</td>
<td>$3.33 \times 10^{-6}$</td>
<td>$2.09 \times 10^{-3}$</td>
</tr>
<tr>
<td>1.6</td>
<td>$1.39 \times 10^{-8}$</td>
<td>$8.72 \times 10^{-6}$</td>
</tr>
<tr>
<td>1.7</td>
<td>$1.47 \times 10^{-7}$</td>
<td>$9.22 \times 10^{-5}$</td>
</tr>
<tr>
<td>1.8</td>
<td>$-1.04 \times 10^{-6}$</td>
<td>$-6.56 \times 10^{-4}$</td>
</tr>
<tr>
<td>1.9</td>
<td>$-7.53 \times 10^{-7}$</td>
<td>$-4.72 \times 10^{-4}$</td>
</tr>
<tr>
<td>2.0</td>
<td>$-2.39 \times 10^{-7}$</td>
<td>$-1.50 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 7.3: KEM energy and energy of $N$-representable density matrix compared to full molecule KS/DFT energy for the linear helium system at various separations, STO-3G basis
Figure 7.1: KEM energy and energy of \( N \)-representable density matrix compared to full molecule KS/DFT energy for the linear helium system at various separations in the STO-3G basis. Taken from Polkosnik and Massa. \(^{51}\)
<table>
<thead>
<tr>
<th>$r$</th>
<th>$E_{\text{KEM}} - E_{\text{full}}$ a.u.</th>
<th>$E_{\text{P}} - E_{\text{full}}$ a.u.</th>
<th>$E_{\text{KEM}} - E_{\text{full}}$ kcal·mol$^{-1}$</th>
<th>$E_{\text{P}} - E_{\text{full}}$ kcal·mol$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>$-1.94 \times 10^{-2}$</td>
<td>$-1.21 \times 10^1$</td>
<td>$-1.21 \times 10^1$</td>
<td>$-1.21 \times 10^1$</td>
</tr>
<tr>
<td>0.5</td>
<td>$-6.36 \times 10^{-2}$</td>
<td>$-3.99 \times 10^1$</td>
<td>$-3.99 \times 10^1$</td>
<td>$-3.99 \times 10^1$</td>
</tr>
<tr>
<td>0.6</td>
<td>$-3.96 \times 10^{-2}$</td>
<td>$-2.48 \times 10^1$</td>
<td>$5.92 \times 10^{-2}$</td>
<td>$3.71 \times 10^1$</td>
</tr>
<tr>
<td>0.7</td>
<td>$-1.88 \times 10^{-2}$</td>
<td>$-1.18 \times 10^1$</td>
<td>$1.30 \times 10^{-2}$</td>
<td>$8.17$</td>
</tr>
<tr>
<td>0.8</td>
<td>$-7.97 \times 10^{-3}$</td>
<td>$-5.00$</td>
<td>$2.87 \times 10^{-3}$</td>
<td>$1.80$</td>
</tr>
<tr>
<td>0.9</td>
<td>$-3.08 \times 10^{-3}$</td>
<td>$-1.93$</td>
<td>$6.76 \times 10^{-4}$</td>
<td>$4.24 \times 10^{-1}$</td>
</tr>
<tr>
<td>1.0</td>
<td>$-1.06 \times 10^{-3}$</td>
<td>$-6.67 \times 10^{-1}$</td>
<td>$2.08 \times 10^{-4}$</td>
<td>$1.31 \times 10^{-1}$</td>
</tr>
<tr>
<td>1.1</td>
<td>$-3.21 \times 10^{-4}$</td>
<td>$-2.01 \times 10^{-1}$</td>
<td>$9.04 \times 10^{-5}$</td>
<td>$5.67 \times 10^{-2}$</td>
</tr>
<tr>
<td>1.2</td>
<td>$-7.38 \times 10^{-5}$</td>
<td>$-4.63 \times 10^{-2}$</td>
<td>$4.56 \times 10^{-5}$</td>
<td>$2.86 \times 10^{-2}$</td>
</tr>
<tr>
<td>1.3</td>
<td>$-3.93 \times 10^{-6}$</td>
<td>$-2.47 \times 10^{-3}$</td>
<td>$2.24 \times 10^{-5}$</td>
<td>$1.41 \times 10^{-2}$</td>
</tr>
<tr>
<td>1.4</td>
<td>$4.09 \times 10^{-6}$</td>
<td>$2.57 \times 10^{-3}$</td>
<td>$1.01 \times 10^{-5}$</td>
<td>$6.35 \times 10^{-3}$</td>
</tr>
<tr>
<td>1.5</td>
<td>$4.20 \times 10^{-6}$</td>
<td>$2.64 \times 10^{-3}$</td>
<td>$4.11 \times 10^{-6}$</td>
<td>$2.58 \times 10^{-3}$</td>
</tr>
<tr>
<td>1.6</td>
<td>$2.16 \times 10^{-8}$</td>
<td>$1.36 \times 10^{-5}$</td>
<td>$1.48 \times 10^{-6}$</td>
<td>$9.31 \times 10^{-4}$</td>
</tr>
<tr>
<td>1.7</td>
<td>$-1.69 \times 10^{-7}$</td>
<td>$1.06 \times 10^{-4}$</td>
<td>$4.67 \times 10^{-7}$</td>
<td>$2.93 \times 10^{-4}$</td>
</tr>
<tr>
<td>1.8</td>
<td>$-7.21 \times 10^{-7}$</td>
<td>$4.52 \times 10^{-4}$</td>
<td>$1.27 \times 10^{-7}$</td>
<td>$8.00 \times 10^{-5}$</td>
</tr>
<tr>
<td>1.9</td>
<td>$-2.53 \times 10^{-7}$</td>
<td>$1.59 \times 10^{-4}$</td>
<td>$3.04 \times 10^{-8}$</td>
<td>$1.91 \times 10^{-5}$</td>
</tr>
<tr>
<td>2.0</td>
<td>$7.48 \times 10^{-7}$</td>
<td>$4.69 \times 10^{-4}$</td>
<td>$4.67 \times 10^{-9}$</td>
<td>$2.93 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

Table 7.4: KEM energy and energy of $N$-representable density matrix compared to full molecule KS/DFT energy for the linear helium system at various separations, 6-31G($d, p$) basis.
Figure 7.2: KEM energy and energy of $N$-representable density matrix compared to full molecule KS/DFT energy the linear helium system at various separations in the 6-31G($d,p$) basis. Taken from Polkosnik and Massa.\textsuperscript{51}
Chapter 8

Discussion and Conclusions

This summarizes the first study of the applications of the ideas of the KEM kernel expansion to full density matrices and the first study that enforces $N$-representability on KEM results. The straightforward KEM expansion for the energy has been demonstrated to deliver accurate results, but these results do not necessarily satisfy the variational theorem because the expansion is not based on a wavefunction for the full system. To address this a kernel expansion for the full system 1-electron density matrix analogous to the KEM energy expansion is defined using augmented kernel density matrices. The Clinton equations are used to impose idempotency and normalization on this matrix obtain a density matrix which is single-determinant $N$-representable. As this $N$-representable density matrix corresponds to a properly normalized antisymmetric wavefunction the energy associated with it satisfies the variational bound in the case of the single-determinant Hartree-Fock theory.

The single-determinant $N$-representable KEM density matrix can be used to obtain the full molecule HF or KS/DFT orbitals. As the matrix delivered by the KEM procedure is a normalized projector, it is factorizable into a product of orbital coefficient matrices.

The methodology was demonstrated on clusters of water molecules consisting of three to twenty water molecules in the context of Hartree-Fock calculations. This was the first application
of KEM to water clusters. For these clusters the simple energy expansion gives accurate energy results, but for some cases the results violate the variational bound. After imposing single-determinant $N$-representability on the initial KEM density matrix the energy of the corresponding $N$-representable density matrix satisfies the variational bound when the KEM energy does not. In addition the $N$-representable density matrix gives energy results that are closer to the full molecule Hartree-Fock energies than the KEM energy expansion in the non-violating cases. This is expected to be true for larger water clusters and for general molecules.

In the context of KS/DFT the KEM density matrix approach delivers a single-determinant $N$-representable full molecule density matrix which gives a more accurate full molecule energy than the KEM energy expansion in the majority of cases. This was shown on a simple proof-of-concept system and on a single water cluster.

The KEM expansion for the electron density, 1-electron density matrix and 2-electron density matrix can be obtained by direct KEM calculation from the coordinates obtained from X-ray crystallography. KEM makes calculations of these quantities on large molecules practicable. In this way the quantum mechanics of large molecules can be extracted from X-ray data. By imposing $N$-representability on the 1-electron density matrix a density matrix and electron density that satisfy the fundamental properties quantum mechanics can be obtained. Using these $N$-representable matrices, X-ray scattering factors can be calculated and used to develop better models for X-ray structure determination.

**Further Work**

The ability to optimize geometries of large molecules would be an extremely useful extension of KEM. A procedure that uses a generalization of the KEM expansion for the analytical gradient is in development.

Further developments for quantum crystallography include the extension of existing KEM code to calculation of X-ray scattering factors from the KEM density matrix. This is a straightforward
extension of the ideas in this thesis.
Appendices
Appendix A

Basis Functions

For the purposes of computation the molecular orbitals are expanded as linear combinations of a set of fixed basis functions.

A Slater-type orbital (STO) centered at $\mathbf{R}$ has the form $^4, ^5$

$$\phi_{\text{STO}}(\mathbf{r} - \mathbf{R}) = \eta (x - X)^a (y - Y)^b (z - Z)^c e^{-\zeta|\mathbf{r} - \mathbf{R}|} \quad (A.1)$$

where $\eta$ is a normalization constant and $\zeta$ is called the orbital exponent. The sum $l = a + b + c$ defines the angular momentum of the orbital, with $l = 0$ corresponding to a s-orbital and $l = 1$ corresponding to a p-orbital. The 1s STO matches the exact 1s solutions for the hydrogen atom. It can be shown that the asymptotic behavior of molecular orbitals matches the exponential falloff of STOs. $^4$ These reasons make Slater type orbitals preferred for quantum chemical calculations, but integral calculations for STOs are computational expensive.

Boys$^{53}$ introduced the Gaussian-type orbital (GTO). A GTO centered at $\mathbf{R}$ has the form

$$\phi_{\text{GTO}}(\mathbf{r} - \mathbf{R}) = \eta (x - X)^a (y - Y)^b (z - Z)^c e^{-\alpha|\mathbf{r} - \mathbf{R}|^2} \quad (A.2)$$

These are efficient for calculation. A common procedure is to fit a linear combination of
fixed GTOs to a STO. An orbital constructed this way is called a *contracted gaussian-type orbital* (cGTO).

\[
\phi_{\text{cGTO}}(\mathbf{r} - \mathbf{R}) = \sum_{p=1}^{L} d_p \phi_{\text{GTO},p}(\alpha_p, \mathbf{r} - \mathbf{R})
\]

(A.3)

where the coefficients \(d\) and exponents \(\alpha\) are chosen to fit a STO function.

One class of basis functions based on cGTOs is denoted by *STO-LG*, with \(L\) denoting the length of the contraction. STO-3G is a commonly used basis set which is computationally inexpensive. It is a *minimal basis set*, in that it requires the least number of functions per atom required to describe the occupied atomic orbitals of that atom.

The minimal basis sets could be improved by including freedom to vary the exponents, however this makes the variational problem nonlinear. A way to include this freedom is to include basis functions of different fixed exponents; the HF procedure will then weigh the short range versus the long range components. In the 6-31G basis set the inner shell functions (1s) for Li-F are a contraction of 6 GTOs while the valence basis functions consist of a set of 3 inner GTO functions and one outer GTO function.

An additional improvement is to add *polarization* functions, or basis functions of higher angular momentum. The basis set 6-31G\((d,p)\) adds a p-type polarization function on hydrogen and d-type functions to heavy (non-hydrogen) atoms.
Appendix B

Water Cluster publication
Single Determinant $N$-Representability and the Kernel Energy Method Applied to Water Clusters*

Walter Polkosnik [a] and Lou Massa [a,b]

The Kernel energy method (KEM) is a quantum chemical calculation method that has been shown to provide accurate energies for large molecules. KEM performs calculations on subsets of a molecule (called kernels) and so the computational difficulty of KEM calculations scales more softly than full molecule methods. Although KEM provides accurate energies those energies are not required to satisfy the variational theorem. In this article, KEM is extended to provide a full molecule single-determinant $N$-representable one-body density matrix. A kernel expansion for the one-body density matrix analogous to the kernel expansion for energy is defined. This matrix is converted to a normalized projector by an algorithm due to Clinton. The resulting single-determinant $N$-representable density matrix maps to a quantum mechanically valid wavefunction which satisfies the variational theorem. The process is demonstrated on clusters of three to twenty water molecules. The resulting energies are more accurate than the straightforward KEM energy results and all violations of the variational theorem are resolved. The $N$-representability studied in this article is applicable to the study of quantum crystallography.

© 2017 Wiley Periodicals, Inc.

**DOI**: 10.1002/jcc.25064

---

**Introduction**

In quantum chemistry, the computational difficulty of solving the Schrödinger equation increases dramatically with the number of basis functions, $N$. For the case of Hartree–Fock theory the difficulty scales as $N^4$. For more accurate methods the difficulty scales as a higher power of $N$. For example, MP2 scales $N^8$ and CI, CISD scales as $N^{10,11}$ The KEM energy method provides softer scaling while maintaining ab initio quality results for the energy.$^{[2–6]}$

In the kernel energy method, the molecule is partitioned into disjoint subsets called kernels. A single kernel is simply a particular choice of a subset of atoms in a molecule. In the case of the water clusters studied in this article, the single kernels are taken to be groups of water molecules. Double kernels consist of the union of the set of atoms in two individual single kernels. The union of all the single kernels is the full molecule. Since the kernels are smaller than the full molecule the computational difficulty scales more softly than in calculations on the full molecule. KEM is particularly appropriate for large molecules and has been found to deliver accurate energy results for biological molecules such as DNA, RNA, and proteins.$^{[7–10]}$

The KEM energy is defined as

$$E_{\text{KEM}} = \sum_{i<j}^K E_{ij} - (K-2) \sum_{i<j}^K E_i$$

where $E_{\text{KEM}}$ is the KEM energy, $E_i$ and $E_{ij}$ are the energies of the single and double kernels, respectively, and $K$ is the total number of single kernels. This expression includes interaction energies between all pairs of kernels, $E_{ij}$, offset by the single kernel energies, $E_i$, which have been overcounted in the double kernel sum. The single and double kernels can be considered as "submolecules" and are, therefore, described by their own molecular Hamiltonians $H_i$ and $H_{ij}$, respectively. $E_i$ is the ordinary expectation value $\langle H_i \rangle$ and similarly $E_{ij}$ is the expectation value $\langle H_{ij} \rangle$. These expectation values are evaluated in the basis function space of the Hamiltonians as usual. In this article, all expectation values correspond to single determinants. The energy $E[R]$ associated with a given density matrix $R$ can be calculated as given explicitly by eq. (20), using the one electron and two electron integrals for a single determinant.

KEM has been demonstrated to give energies close to full molecule ab initio energy calculations. But there is no mathematical requirement that the energy given by KEM satisfies the requirements for $N$-representability, which include the variational theorem. A quantum mechanical requirement for energy calculations arising from a Hamiltonian expectation value, $\langle \psi | H | \psi \rangle$, calculated using an antisymmetric $N$-body wavefunction $\psi$, is that they must satisfy $E \geq E_0$ where $E_0$ is the exact ground state energy for that Hamiltonian.$^{[11]}$

Our goal in this article is to investigate such violations of the variational principle, and, therefore, of $N$-representability, and to propose an extension of KEM so that the variational
principle and \( N \)-representability are satisfied. We will introduce a generalization of the kernel energy expansion to density matrices and then impose single-determinant \( N \)-representability on this matrix using the Clinton algorithm. The resulting density matrix will map to a single normalized Slater determinant wavefunction which satisfies the variational theorem. The one-body density matrix allows calculation of the expectation value of any one-body operator while the trace of the density matrix represents the electron density. In particular, the density allows calculation of coherent X-ray scattering structure factors.

**Discussion**

The one-body density matrix for an \( N \)-electron molecule is defined as

\[
\rho_1(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}_2 \cdots d\mathbf{r}_n \Psi^\dagger(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_n)\Psi(\mathbf{r}', \mathbf{r}_2, \ldots, \mathbf{r}_n) \tag{2}
\]

The one-body density matrix is a normalized projector

\[
\int d\mathbf{r} \rho_1(\mathbf{r}, \mathbf{r}) = N \tag{3}
\]

\[
\int d\mathbf{r}' \rho_1(\mathbf{r}, \mathbf{r}') \rho_1(\mathbf{r}', \mathbf{r}) = \rho_1(\mathbf{r}, \mathbf{r}') \tag{4}
\]

In the Hartree–Fock method in particular the set of molecular orbitals \( \phi \) are expanded in a basis \( \Psi \), with \( \phi = \mathbf{C}\Psi \). In the case of doubly occupied orbitals the density matrix reduces to

\[
\rho_1(\mathbf{r}, \mathbf{r}') = 2tr(\phi(\mathbf{r}')\phi^\dagger(\mathbf{r})) = 2tr(\mathbf{C}\Psi(\mathbf{r}')^\dagger\Psi^\dagger(\mathbf{r})) = 2tr(\mathbf{R}\Psi(\mathbf{r}')^\dagger\Psi^\dagger(\mathbf{r}) \tag{5}
\]

The matrix \( \mathbf{R} = \mathbf{C}^\dagger\mathbf{C} \) is the density matrix expressed in the \( \Psi \) basis.

We define the corresponding KEM version of this density matrix by analogy to the KEM energy expression.

\[
\mathbf{R}_{\text{KEM}} = \sum_{i<j}^K \mathbf{R}_{ij} - (K-2) \sum_i^K \mathbf{R}_i \tag{6}
\]

The matrices \( \mathbf{R}_{ij} \) and \( \mathbf{R}_i \) are constructed from elements of the single and double kernel density matrices, here denoted by \( \mathbf{r}_i \) and \( \mathbf{r}_ij \). The matrices \( \mathbf{r}_i \) and \( \mathbf{r}_ij \) are defined identically as in eq. (5) but are constructed in the bases of individual single and double kernels and as such are of smaller dimension than the full molecule density matrix. Since each kernel density matrix operates only on a subset of the full molecule basis each kernel density matrix must be adapted to the full molecule basis to represent the contribution of that kernel to the full molecule. This is done by augmenting the original kernel matrices by placing the elements of the kernel matrices in their appropriate locations in the full molecule matrix with all other elements set to zero. For a single kernel density matrix denoted by \( \mathbf{r}_{ij} \), the form of the corresponding augmented matrix is

\[
\mathbf{R}_{ij} = \begin{pmatrix}
0 & 0 & 0 \\
0 & \mathbf{r}_{ij} & 0 \\
0 & 0 & 0
\end{pmatrix} \tag{7}
\]

Here, \( \mathbf{r}_{ij} \) is the entire single kernel density matrix set as a block into the appropriate position in a matrix which is the size of the full molecule matrix.

With \( n_i \) defined as the basis set size of kernel \( i \), the augmented matrix can be written explicitly in terms of basis indices. A single kernel augmented matrix \( \mathbf{r}_{ij} \) is

\[
\mathbf{R}_{ij} = \begin{pmatrix}
(\mathbf{r}_{ij})_{\mu\mu} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & (\mathbf{r}_{ij})_{\nu\nu} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0}
\end{pmatrix} \tag{8}
\]

For a double kernel density matrix \( \mathbf{r}_{ij} \), the corresponding augmented matrix is of the form

\[
\mathbf{R}_{ij} = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
0 & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{pmatrix} \tag{10}
\]

In eq. (10) subscripts denote kernels while superscripts denote blocks of the density matrix for that kernel. The matrix \( (\mathbf{r}_{ij})_{\mu\mu} \) is the block from the double kernel density matrix that corresponds to basis function products for atoms only in kernel \( i \) and \( (\mathbf{r}_{ij})_{\nu\nu} \) is the block referring to basis function products for atoms only in kernel \( j \). The off-diagonal blocks \( (\mathbf{r}_{ij})^{\mu\nu} \) and \( (\mathbf{r}_{ij})^{\nu\mu} \) are elements of the double kernel density matrix which multiply basis function products from atoms in kernels \( i \) and \( j \). The remaining elements of the matrix are set to zero.

The blocks of the augmented double kernel matrix are defined through the following four equations.

\[
\mathbf{R}_{ij}^{\mu\mu} = \begin{pmatrix}
(\mathbf{r}_{ij})_{\mu\mu} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & (\mathbf{r}_{ij})_{\nu\nu} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0}
\end{pmatrix} \tag{11}
\]

\[
\mathbf{R}_{ij}^{\mu\nu} = \begin{pmatrix}
(\mathbf{r}_{ij})_{\mu\nu} & \mathbf{0} \tag{12}
\end{pmatrix}
\]

\[
\mathbf{R}_{ij}^{\nu\mu} = \begin{pmatrix}
(\mathbf{r}_{ij})_{\nu\mu} \tag{13}
\end{pmatrix}
\]

\[
\mathbf{R}_{ij}^{\nu\nu} = \begin{pmatrix}
(\mathbf{r}_{ij})_{\nu\nu} & \mathbf{0} \tag{14}
\end{pmatrix}
\]

The density matrix \( \mathbf{P} \) is defined in terms of the \( \mathbf{R} \) density matrix through the square root of the overlap matrix \( \mathbf{S} \). The square root of \( \mathbf{S} \) is used to obtain a \( \mathbf{P} \) which is symmetric.

\[
\mathbf{S} = \int d\mathbf{r} \Psi(\mathbf{r})\Psi^\dagger(\mathbf{r}) \tag{15}
\]

\[
\mathbf{P} = \mathbf{S}^{1/2} \mathbf{R} \mathbf{S}^{1/2} \tag{16}
\]

The normalization and the projector property must be enforced on the \( \mathbf{P} \) matrix as follows from eqs. (3), (5), and (15). The overlap matrix \( \mathbf{S} \) in the full molecule basis is obtained from the double and single kernel matrices through a process similar to the construction of the augmented density matrix.
In the case of \( S \), matrix elements from the kernel overlap matrices are simply placed in appropriate positions.

The \( P_{\text{KEM}} \) matrix corresponding to \( R_{\text{KEM}} \) is normalized by construction, but is not necessarily a projector and so does not satisfy the conditions for \( N \)-representability. Only \( N \)-representable one-body density matrices correspond to quantum mechanically valid \( N \)-electron antisymmetric wavefunctions. This correspondence is illustrated in Figure 1.

Single determinant \( N \)-representability will be imposed by the Clinton algorithm. Starting from a \( P_{\text{KEM}} \) which is not a normalized projector the Clinton algorithm produces a density matrix which is a normalized projector. This guarantees single-determinant \( N \)-representability. The resulting projector determines a set of orbitals which define a single Slater determinant, which guarantees its corresponding energy satisfies the variational principle. In terms of \( P \) the normalization condition is written \( 2trP=N \) and the projector condition is \( P^2=P \). The projector property and normalization are enforced on a density matrix, \( P \) by extremizing the squared norm of the difference between \( P^2 \) and \( P \) subject to a normalization constraint. The quantity extremized is

\[
\text{tr}(P^2 - P) - \lambda(\text{tr}P - N/2)
\]

where \( \text{tr}(P^2 - P) \) is the square of the Frobenius norm of the difference between \( P^2 \) and \( P \) and \( \lambda \) is a Lagrange multiplier that enforces the normalization constraint, \( 2\text{tr}P=N \), for doubly occupied orbitals.

Extremizing eq. (17) gives an equation that can be solved using a fixed-point iteration approach that delivers a normalized projector starting from an arbitrary matrix. In our case, the initial matrix used to start the iterations is \( P_0=P_{\text{KEM}} \). The form of the iterative equation is as follows.

\[
P_{n+1} = 3P_n^2 - 2P_n^3 + \lambda_n \mathbb{1}
\]

\[
\lambda_n = \frac{N/2 - \text{tr}(3P_n^2 + 2P_n^3)}{\text{tr} \mathbb{1}}
\]

The energy \( E[R] \) associated with a given density matrix can be calculated using the one electron and two electron integrals for the full molecule. These integrals depend only on the basis functions and the atomic coordinates on which they are centered and are independent of the density matrix for a given system. The expression for \( E[R] \) is

\[
E[R] = V^{\text{noc}} + \sum_{\mu \nu} T_{\mu \nu} R_{\mu \nu} + \frac{1}{2} \sum_{\mu \nu \rho \sigma} V^{\text{e-e}}_{\mu \nu \rho \sigma} (R_{\mu \nu} R_{\rho \sigma} - \frac{1}{2} R_{\mu \nu} R_{\rho \sigma}).
\]

The \( V^{\text{noc}} \) term represents the classical nuclear–nuclear potential energy. The integrals for the electronic kinetic energy and electron–nuclear potential are \( T_{\mu \nu} \) and \( V_{\mu \nu} \), respectively, while the integrals corresponding to the electron-electron potential are \( V^{\text{e-e}}_{\mu \nu \rho \sigma} \). The expression, \( R_{\mu \nu} R_{\rho \sigma} - \frac{1}{2} R_{\mu \nu} R_{\rho \sigma} \), in the last term of this equation is the matrix representative of the two-body density matrix for a single determinant.

Our investigation of KEM \( N \)-representability employs a set of water clusters. Water clusters have not been studied previously with KEM. Water clusters are ubiquitous in our environment and are influential in determining the geometry of proteins, DNA, and many other types of biological structures. They are especially interesting to study by means of KEM in the context of such large systems. We investigate whether KEM energy calculations on these clusters satisfy the variational theorem. In those cases in which it is not satisfied we impose \( N \)-representability and note the effect on the resulting energy. Imposing \( N \)-representability on the union of fragments as done in this article was prefigured in a work by Hernandez and Bader, but their fragments were Bader-atoms defined by the Bader zero flux condition \( \nabla \rho \cdot \mathbf{n} = 0 \) and did not invoke any conditions related to energy as occurs with the KEM considerations of this article.

**Results**

We investigate imposing single-determinant \( N \)-representability in the context of water clusters. We compute KEM energies of a set water clusters to assess their accuracy, whether the energies satisfy the variational theorem and the effect on the energies due to imposing \( N \)-representability on the KEM density matrix. We will demonstrate that KEM gives accurate energies which, in some cases, violate the variational principle and that imposing \( N \)-representability gives accurate energies that do not violate the variational theorem.

The structures used in these calculations were obtained from Ref. [16]. Each structure is at an energy minimized geometry. A total of nineteen water clusters were studied, ranging in size from three to twenty molecules. The six molecule cluster consisted of two structures, a cage and prism arrangement. Each cluster was partitioned into three kernels of roughly equal size by cutting across hydrogen bonds. All KEM calculations were done using restricted Hartree–Fock theory, which is a single-determinant method. Calculations were performed in the Gaussian program using a 6–31G basis set. Energies, matrix elements, and integrals were all obtained from Gaussian output.
Results of KEM energy calculations on are shown in Table 1. The difference between the KEM energy and the full molecule energies ranges from -0.41 kcal mol$^{-1}$ and 7.39 kcal mol$^{-1}$. Nearly all of the KEM energies, 15 out of 19, satisfy the variational theorem $E_{\text{KEM}} \geq E_{\text{full}}$. The four cases which violate the variational theorem are $N_{\text{water}} = 10, 12, 14, 15$. The KEM energies occasionally depart substantially from the full molecule Hartree–Fock energies, as occurs in Table 1 for the cases $N = 8$ and $N = 16$. While there is not an obvious reason to explain these larger than average deviations it is satisfying to note how much better these results become once $N$-representability is imposed, as indicated in Table 2.

To investigate the effect of the Clinton algorithm on the KEM energies $R_{\text{KEM}}$ was constructed for each cluster and the Clinton algorithm applied to obtain a normalized projector. The criterion for convergence of the Clinton algorithm was $\text{tr}(P^2 - P)^2 \leq 10^{-18}$. To obtain the energy of $P_{\text{projector}}$ eq. (20) was used after transforming back to the $R$ basis.

Results are summarized in Table 2. The deviation from the full molecules energy range from 0.20 kcal mol$^{-1}$ to 1.85 kcal mol$^{-1}$. Note that in each case the energy of the normalized projector is closer to the full molecule energy than $E_{\text{KEM}}$. Most notably, in all the cases where there was a violation of the variational bound the Clinton algorithm removes the violation.

Details of the Clinton algorithm iterations are presented in Tables (3–7). Iteration number $n = 0$ corresponds to the initial $P_{\text{KEM}}$ matrix obtained from the augmented kernel density matrix sum while the last iteration corresponds to $P_{\text{projector}}$. The energy associated with $P_n$ at each step of the Clinton algorithm is calculated using eq. (20).

The $N_{\text{water}} = 20$ case is presented in Table 3. This is a case that does not violate the variational bound. The measure
deviation from a projector converges to below tolerance in only three steps. The general behavior of all the other nonviolating cases is similar. Each converges in a few steps and each yields an energy closer to the full molecule energy than $E_{\text{full}}$.

The details for the four cases that violate the variational theorem are presented in Tables (4–7). As the iterations progress the $P$ matrix quickly converges to a normalized projector. In each of these cases where there was violation of the variational theorem the violation was removed by the Clinton algorithm.

To further test the accuracy of KEM and possible violations of the variational theorem each calculation was done on the original structure and on the structure scaled down by a factor of 0.55. Structures scaled by smaller factors were attempted but failed to converge in Gaussian due to small interatomic distances. This left 190 structures for which $E_{\text{full}}$, $E_{\text{KEM}}$, and $P_{\text{KEM}}$ were calculated. The Clinton algorithm was applied to each $P_{\text{KEM}}$ and $E[P_{\text{projector}}]$ was obtained.

Comparing $E_{\text{KEM}}$ to $E_{\text{full}}$, the difference between the KEM energy and the full molecule energy ranged from $1.4 \times 10^{-4}$ a.u. to $4.2 \times 10^{-2}$ a.u. The KEM results are accurate even for systems scaled down by a factor of 0.55.

There were 80 violations of the variational bound out of a total 190 structures. The Clinton algorithm was applied to each initial $P_{\text{KEM}}$ to obtain a normalized projector. In each case the Clinton algorithm converged to a normalized projector and in each case $E[P_{\text{projector}}]$ was greater than $E_{\text{full}}$. Every violation of the variational bound was lifted on refining $P_{\text{KEM}}$ using the Clinton algorithm. The final energy differences between $E[P_{\text{projector}}]$ and $E_{\text{full}}$ ranged from $3.2 \times 10^{-4}$ a.u. to $8.2 \times 10^{-2}$ a.u. The variational theorem is always satisfied, as it must be, for $E[P_{\text{projector}}]$. This is the principal result of this article.

### Conclusions

This is the first study of $N$-representability in the context of KEM and the first study of the KEM approximation applied to clusters of water molecules. Clusters of water molecules are important to study with KEM because of the important role they play in biological systems. We find that KEM delivers energies close to the full molecule energy in all the clusters studied. This is expected to be true for KEM calculations of larger water clusters as well as larger molecules in general. While KEM gives good energy results there is no mathematical requirement that KEM energies satisfy the variational principle and in many of the calculations here the variational principle was violated. To address these violations of the variational theorem we have shown how to construct a KEM density matrix and how to obtain a normalized projector from this initial matrix using the Clinton algorithm. The resulting matrix is single-determinant $N$-representable and, therefore, satisfies the variational theorem. After imposing $N$-representability the energy corresponding to the normalized projector density matrix is in fact closer to the Hartree–Fock energy of the full molecule.

The significance of $N$-representability for the study of quantum crystallography\textsuperscript{[18,19]} is as follows. Analogous to what has been done here with water clusters, one may also calculate the quantum mechanics of the kernels composing a crystal. One can then put the crystal back into its entirety by properly summing the kernels, at the same time attaching Debye–Waller factors from the experiment to the basis functions. The resulting density can then be used to calculate the X-ray scattering factors. Such a quantum crystallography/kernel energy method (QCr/KEM) procedure would allow the density, 1-density matrix, and 2-density matrix to be extracted in KEM form. Crystallography provides the atomic coordinates and their fluctuations, while KEM provides the density matrices of the kernels inherent in the X-ray coordinates. In such fashion, we argue it is possible to extract true quantum mechanics from the crystallography experiment. This is important to quantum crystallography in as much as it means that true quantum mechanics including its $N$-representability requirement can be extracted from X-ray data.

### Acknowledgments

We thank Hiroko Ajiki for discussions regarding water clusters. We are grateful to Gaussian, Inc. for assistance with the use of their software. We thank CUNY for use of computer resources.

**Keywords:** soft scaling quantum chemical methods · density matrix · $N$-representability · water clusters · quantum crystallography

To cite this article: W. Polkosnik, L. Massa J. Comput. Chem. 2017, DOI: 10.1002/jcc.25064

---

**Table 6.** Clinton algorithm applied to obtain $P_{\text{projector}}$ for $N_{\text{water}}=14$.

<table>
<thead>
<tr>
<th>n</th>
<th>$trP_n$</th>
<th>$tr((P_n^2-P_n)^2)$</th>
<th>$E[P_n]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>70.000000</td>
<td>$1.12 \times 10^{-3}$</td>
<td>1.800000</td>
</tr>
<tr>
<td>1</td>
<td>70.000000</td>
<td>$1.00 \times 10^{-6}$</td>
<td>1.530000</td>
</tr>
<tr>
<td>2</td>
<td>70.000000</td>
<td>$1.55 \times 10^{-12}$</td>
<td>2.400000</td>
</tr>
<tr>
<td>3</td>
<td>70.000000</td>
<td>$8.89 \times 10^{-24}$</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

In this case, $E_{\text{KEM}}=-1064.057735$ a.u., $E_{\text{full}}=-1064.056877$ a.u., $E_{\text{full}}-E_{\text{KEM}}=-8.58 \times 10^{-4}$ a.u., and $E[P_{\text{projector}}]=-1064.055613$ a.u., $E[P_{\text{projector}}]-E_{\text{full}}=1.26 \times 10^{-3}$ a.u.

**Table 7.** Clinton algorithm applied to obtain $P_{\text{projector}}$ for $N_{\text{water}}=15$.

<table>
<thead>
<tr>
<th>n</th>
<th>$trP_n$</th>
<th>$tr((P_n^2-P_n)^2)$</th>
<th>$E[P_n]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>75.000000</td>
<td>$1.81 \times 10^{-3}$</td>
<td>2.730000</td>
</tr>
<tr>
<td>1</td>
<td>75.000000</td>
<td>$2.88 \times 10^{-6}$</td>
<td>4.150000</td>
</tr>
<tr>
<td>2</td>
<td>75.000000</td>
<td>$1.52 \times 10^{-11}$</td>
<td>2.280000</td>
</tr>
<tr>
<td>3</td>
<td>75.000000</td>
<td>$8.47 \times 10^{-22}$</td>
<td>1.460000</td>
</tr>
</tbody>
</table>

In this case, $E_{\text{KEM}}=-1140.098043$ a.u., $E_{\text{full}}=-1140.096400$ a.u., $E_{\text{full}}-E_{\text{KEM}}=-1.64 \times 10^{-3}$ a.u., and $E[P_{\text{projector}}]=-1140.094095$ a.u., $E[P_{\text{projector}}]-E_{\text{full}}=2.31 \times 10^{-3}$ a.u.

---


\[2\] L. Huang, C. Matta, L. Massa, Struct. Chem. 2015, 26, 1433.


\[4\] L. Huang, L. Massa, C. F. Matta, Carbon 2014, 76, 310.
Appendix C

KS/DFT publication
Kohn-Sham Density Matrix and the Kernel Energy Method

POLKOSNIK Walter$^{1,*}$, MASSA Lou$^{2,3}$

$^1$The Graduate Center of the City University of New York, Physics Department, 365 5th Avenue, New York, NY 10016, USA.
$^2$Hunter College of the City University of New York, Chemistry Department, 695 Park Avenue, New York, NY 10065, USA.
$^3$The Graduate Center of the City University of New York, Physics and Chemistry Departments, 365 5th Avenue, New York, NY 10016, USA.

Abstract: The kernel energy method (KEM) has been shown to provide fast and accurate molecular energy calculations for molecules at their equilibrium geometries. KEM breaks a molecule into smaller subsets, called kernels, for the purposes of calculation. The results from the kernels are summed according to an expression characteristic of KEM to obtain the full molecule energy. A generalization of the kernel expansion to density matrices provides the full molecule density matrix and orbitals. In this study, the kernel expansion for the density matrix is examined in the context of density functional theory (DFT) Kohn-Sham (KS) calculations. A kernel expansion for the one-body density matrix analogous to the kernel expansion for energy is defined, and is then converted into a normalized projector by using the Clinton algorithm. Such normalized projectors are factorizable into linear combination of atomic orbitals (LCAO) matrices that deliver full-molecule Kohn-Sham molecular orbitals in the atomic orbital basis. Both straightforward KEM energies and energies from a normalized, idempotent density matrix obtained from a density matrix kernel expansion to which the Clinton algorithm has been applied are compared to reference energies obtained from calculations on the full system without any kernel expansion. Calculations were performed both for a simple proof-of-concept system consisting of three atoms in a linear configuration and for a water cluster consisting of twelve water molecules. In the case of the proof-of-concept system, calculations were performed using the STO-3G and 6-31G(d, p) bases over a range of atomic separations, some very far from equilibrium. The water cluster was calculated in the 6-31G(d, p) basis at an equilibrium geometry. The normalized projector density energies are more accurate than the straightforward KEM energy result in nearly all cases. In the case of the water cluster, the energy of the normalized projector is approximately four times more accurate than the straightforward KEM energy result. The KS density matrices of this study are applicable to quantum crystallography.

Key Words: Kohn Sham density matrix; Kernel energy method; N-representability; Quantum crystallography; Water cluster

1 Introduction

In quantum chemistry the computational difficulty of solving the Schrödinger equation increases dramatically with the number of basis functions. The computational difficulty of the kernel energy method scales more softly and has been shown to provide accurate energy results over a wide range of systems at their equilibrium geometries. In the kernel energy method the molecule is cut into subsets called kernels. The KEM energy is defined as

$$E_{\text{KEM}} = \sum_{i<j}^K E_{ij} - (K^2 - 2K) \sum_i^K E_i$$

(1)

where $E_{\text{KEM}}$ is the full molecule energy, $E_i$ and $E_{ij}$ are the energies of single and double kernels respectively, and $K$ is the total number of single kernels. Double kernels consist of the union of atoms in the single kernels. While KEM is known to generally give energies close to full molecule ab-initio energies.
it does not deliver the density matrix of the full molecule, nor the corresponding orbitals.

In this paper we use a generalization of KEM that provides the full molecule KS orbitals and KS density matrix from KS/DFT calculations on individual kernels. The kernel expansion of the density matrix is defined for the full molecule and then the conditions of idempotency and normalization are imposed on it by an algorithm due to Clinton. The resulting density matrix, because it is factorizable into a product of LCAO coefficient matrices, maps to a set of full molecule KS orbitals. Those orbitals and density allow calculation of all KS expectation values. In particular, the KS density allows calculation of coherent X-ray scattering structure factors of crystallized molecules.

2 Discussion

Here we review the fundamental ideas related to the use of KEM to obtain a normalized idempotent density matrix from KS/DFT results.\(^{10}\)

In the Kohn Sham case we consider here, the set of molecular orbitals \(\phi\) are expanded in an atomic orbital basis \(\psi\) with \(\phi = C\psi\). For doubly occupied orbitals the density matrix is defined by the following equation.

\[
\rho_1(r, r') = 2 \text{tr} \phi(r')\phi(r) = 2 \text{tr} C\psi(r')\psi(r)C^\dagger = 2 \text{tr} R\psi(r')\psi(r) \tag{2}
\]

The matrix \(R = C^\dagger C\) is the density matrix expressed in the \(\psi\) basis. The KEM version of this density matrix is defined by analogy to the KEM energy expression.

\[
R_{\text{KEM}} = \sum_{i<j} R_{ij} - (K-2) \sum_i R_i \tag{3}
\]

The matrices \(R_i\) and \(R_{ij}\) are constructed from the individual single and double kernel density matrices to represent their contribution to the full molecule density. The KEM density matrix \(R_{\text{KEM}}\) contains contributions from all the single and double kernels that constitute the full molecule. The union of the single kernels represents all the atoms in the full molecule. The double kernels include the interactions between all pairs of single kernels. In this way, \(R_{\text{KEM}}\) is an approximation to the global density matrix for the full system. While \(R_{\text{KEM}}\) is not \(N\)-representable, \(N\)-representability will be imposed by the Clinton algorithm as described below to provide a \(N\)-representable global density matrix for the full system.

Each kernel density matrix must be adapted to the full molecule basis to represent the contribution of that kernel to the full molecule. This is done by augmenting the kernel matrices as we now indicate.

For a single kernel density matrix, denoted by \(r_i\), the form of the corresponding augmented matrix is

\[
R_i = \begin{pmatrix}
0 & 0 & 0 \\
0 & r_i & 0 \\
0 & 0 & 0
\end{pmatrix} \tag{4}
\]

where \(r_i\) is the entire single kernel density matrix set as a block into the appropriate position in a matrix which is the size of the full molecule matrix.

For a double kernel density matrix, \(r_{ij}\), the corresponding augmented matrix is of the form

\[
R_{ij} = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & (r_{ij})^{ii} & 0 & (r_{ij})^{ij} & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & (r_{ij})^{ji} & 0 & (r_{ij})^{jj} & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix} \tag{5}
\]

where \((r_{ij})^{ii}\) is the block from the double kernel density matrix that corresponds to basis function products for atoms only in kernel \(i\) and \((r_{ij})^{jj}\) is the block referring to basis function products for atoms only in kernel \(j\). The off-diagonal blocks \((r_{ij})^{ij}\) and \((r_{ij})^{ji}\) are elements of the double kernel density matrix which multiply basis function products from atoms in kernels \(i\) and \(j\). The remaining elements of the matrix are set to zero.

The density matrix \(P\) is defined in terms of the \(R\) density matrix through the square root of the overlap matrix \(S\).

\[
P = S^{\frac{1}{2}} R S^{\frac{1}{2}} \tag{6}
\]

The overlap matrix \(S\) in the full molecule basis is obtained from the single and double kernel matrices through a process similar to the construction of the augmented kernel density matrices. In the case of \(S\), however, the matrix elements from the kernel overlap matrices are simply placed in their appropriate positions.

The normalized projector condition will be imposed by the Clinton equations.\(^{11}\) In terms of \(P\), the normalization condition for double occupied orbitals is \(\text{tr} P = N/2\) and the projector condition is \(P^2 = P\). The Clinton equations are obtained from extremizing

\[
\text{tr}(P^2 - P)^2 - \lambda (\text{tr} P - N/2) \tag{7}
\]

where \(\text{tr}(P^2 - P)^2 = \|P^2 - P\|^2\) is the square of the Frobenius norm of the difference between \(P^2\) and \(P\) and \(\lambda\) is a Lagrange multiplier that enforces the normalization constraint. The Clinton equations are

\[
P_n+1 = 3P_n^2 - 2P_n^3 + \lambda_n I \tag{8}
\]

\[
\lambda_n = \frac{[N/2 - \text{tr}(3P_n^2 + 2P_n^3)]/\text{tr} I} \tag{9}
\]

As the iteration number \(n\) increases, the Lagrangian multiplier \(\lambda_n\) used to impose normalization, goes to zero, and the matrix \(P\) goes to a normalized projector. The resulting projector determines a set of full molecule KS orbitals which deliver the KS electron density.\(^{12}\).
3 Results

To illustrate the formalism discussed in the previous section we study a simple proof of concept system consisting of three helium atoms in a linear configuration symmetric about a plane through the middle atom. Calculations for a variety of distances separating the atoms were done to explore the accuracy of both the KEM energy and the energy associated with the normalized projector KS/KEM density matrix. The range of the distances separating the atoms was taken to be 0.4 \( R_{vdW} < R < 2.0 \) \( R_{vdW} \), where \( R_{vdW} = 1.4 \ \text{Å} \) (1 Å = 0.1 nm) is the van der Waals radius of helium. At the minimum separation the atoms are heavily clashing. At distances smaller than 0.4 \( R_{vdW} \) KS/DFT calculations on the double kernels are not possible due to close contacts. The greatest separation used was twice the van der Waals radius of helium. At this distance each atom is essentially independent of the others and results are not expected to change appreciably at larger distances. The three single kernels were taken to be the individual helium atoms, thus the double kernels consist of the three pairings of these three single kernels.

The calculations for the proof of concept system are all based upon the chemical model KS/DFT B3LYP, using STO-3G and 6-31G(d,p) basis sets. Results for the energy and density matrix elements were extracted from the Gaussian program. The KS energy \( E[P] \) associated with a given density matrix was also calculated using Gaussian.¹³

The numerical results for \( E_{\text{KEM}} \) over the range of distances considered are shown in Table 1 for the STO-3G basis and in Table 3 for the 6-31G(d,p) basis. The energies obtained from the KEM density matrix after imposing normalization and idempotency using the Clinton algorithm are presented in Table 2 for the STO-3G basis and in Table 4 for the 6-31G(d,p) basis. In all these tables \( E_{\text{Full}} \) is the energy of the full system, which is provided as a reference for comparison to the KEM results.

For large \( R \), i.e. \( R/R_{vdW} > 1 \), Tables 1–4 show that the direct KEM energy, \( E_{\text{KEM}} \), and the KEM projector energy, \( E[P_{\text{projector}}] \), are close to the full molecule energy in both the STO-3G and 6-31G(d,p) basis. As the distance of separation \( R \) decreases, however, the direct KEM energy values in the STO-3G basis (Table 1) begin to steadily diverge from the full molecule results while the normalized projector energy (Table 2) is accurate at all distances. At the minimum separation distance of 0.4 \( R_{vdW} \) the energy difference \( E_{\text{KEM}} - E_{\text{Full}} \) has risen to a very large magnitude of error, \(-137 \text{ kcal/mol}^{\text{1}}, \) while \( E[P_{\text{projector}}] - E_{\text{Full}} = 1.84 \times 10^{-5} \text{ kcal/mol}^{-1} \), a very small error. Even at large separations the projector energy is much closer to the full molecule energy.

For calculations on the proof of concept system in the 6-31G(d,p) basis the KEM energy and the KEM projector energy both track the full molecule energy for \( R/R_{vdW} > 1 \). In contrast with the results for the STO-3G basis, the straightforward KEM energy and the projector energy both diverge from the full molecule energy as the separation between the atoms decreases. Also in contrast with the results in the STO-3G basis, in the cases \( R/R_{vdW} = 1.8, 1.7, 1.4, 1.3, 0.6 \), the straightforward KEM energy is more accurate than the projector energy. The projector energy in the majority of cases, however, is more accurate than the KEM energy.

The trends we have observed for the proof of concept system in the STO-3G basis are also presented in Fig. 1. In Fig. 1, for interatomic distances \( R/R_{vdW} > 1 \) the KEM energy, \( E_{\text{KEM}} \) clearly closely tracks the full molecule energy, \( E_{\text{Full}} \) while for \( R/R_{vdW} < 1 \), \( E_{\text{KEM}} \) diverges from \( E_{\text{Full}} \). The energy of the projector, \( E[P_{\text{projector}}] \), is accurate over the whole range of atomic separations. The energy obtained from the KS/KEM density matrix \( E[P_{\text{projector}}] \) is very close to the full system energy even

<table>
<thead>
<tr>
<th>Table 1 KEM energy results, ( E_{\text{KEM}} ), compared to full molecule energies ( E_{\text{Full}} ), both calculated in the STO-3G basis, for the linear helium system over a range of atomic separations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R/R_{vdW} ) &amp; ( E_{\text{KEM}} ) (a.u.) &amp; ( E_{\text{Full}} ) (a.u.) &amp; ( E_{\text{KEM}} - E_{\text{Full}} ) (kcal/mol(^{-1}))</td>
</tr>
<tr>
<td>0.4 &amp; -5.98669314 &amp; -5.76866548 &amp; -2.81 \times 10^{-5} -1.37 \times 10^{-4}</td>
</tr>
<tr>
<td>0.5 &amp; -7.22786368 &amp; -7.13655432 &amp; -9.13 \times 10^{-5} -5.73 \times 10^{-4}</td>
</tr>
<tr>
<td>0.6 &amp; -7.85016146 &amp; -7.81943262 &amp; -3.82 \times 10^{-5} -2.40 \times 10^{-4}</td>
</tr>
<tr>
<td>0.7 &amp; -8.17778919 &amp; -8.16251286 &amp; -1.53 \times 10^{-5} -9.59</td>
</tr>
<tr>
<td>0.8 &amp; -8.35417031 &amp; -8.33585764 &amp; -5.78 \times 10^{-5} -3.63</td>
</tr>
<tr>
<td>0.9 &amp; -8.49812227 &amp; -8.47775812 &amp; -2.04 \times 10^{-4} -1.28</td>
</tr>
<tr>
<td>1.0 &amp; -8.50163221 &amp; -8.50097018 &amp; -6.62 \times 10^{-6} -4.15 \times 10^{-5}</td>
</tr>
<tr>
<td>1.1 &amp; -8.52932035 &amp; -8.52912163 &amp; -1.99 \times 10^{-4} -1.25 \times 10^{-4}</td>
</tr>
<tr>
<td>1.2 &amp; -8.54373793 &amp; -8.54369095 &amp; -4.70 \times 10^{-5} -2.95 \times 10^{-4}</td>
</tr>
<tr>
<td>1.3 &amp; -8.55109639 &amp; -8.55109395 &amp; -2.44 \times 10^{-6} -1.53 \times 10^{-4}</td>
</tr>
<tr>
<td>1.4 &amp; -8.55480498 &amp; -8.55480498 &amp; -2.47 \times 10^{-6} -1.55 \times 10^{-3}</td>
</tr>
<tr>
<td>1.5 &amp; -8.55660335 &amp; -8.55666668 &amp; -3.33 \times 10^{-6} -2.09 \times 10^{-3}</td>
</tr>
<tr>
<td>1.6 &amp; -8.55743727 &amp; -8.55743728 &amp; -1.39 \times 10^{-6} -8.72 \times 10^{-6}</td>
</tr>
<tr>
<td>1.7 &amp; -8.55781516 &amp; -8.55781531 &amp; -1.47 \times 10^{-7} -9.22 \times 10^{-5}</td>
</tr>
<tr>
<td>1.8 &amp; -8.55797273 &amp; -8.55797168 &amp; -1.04 \times 10^{-6} -6.65 \times 10^{-4}</td>
</tr>
<tr>
<td>1.9 &amp; -8.55805781 &amp; -8.55805706 &amp; -7.53 \times 10^{-7} -4.72 \times 10^{-4}</td>
</tr>
<tr>
<td>2.0 &amp; -8.55808689 &amp; -8.55808665 &amp; -2.39 \times 10^{-7} -1.50 \times 10^{-4}</td>
</tr>
</tbody>
</table>

1 cal = 4.1868 J
Table 3 KEM energy results, $E_{\text{KEM}}$, compared to full molecule energies, $E_{\text{full}}$, both calculated in the 6-31G(d, p) basis, for the linear helium system over a range of atomic separations.

<table>
<thead>
<tr>
<th>$R/R_{\text{csw}}$</th>
<th>$E_{\text{KEM}}$ (a.u.)</th>
<th>$E_{\text{full}}$ (a.u.)</th>
<th>$E_{\text{KEM}} - E_{\text{full}}$ (kcal-mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>$-6.956944817$</td>
<td>$-6.93758478$</td>
<td>$-1.14\times10^{-4} - 1.21\times10^{-3}$</td>
</tr>
<tr>
<td>0.5</td>
<td>$-7.31277299$</td>
<td>$-7.66770987$</td>
<td>$-3.36\times10^{-2} - 3.99\times10^{-3}$</td>
</tr>
<tr>
<td>0.6</td>
<td>$-8.16007280$</td>
<td>$-8.21058176$</td>
<td>$-3.96\times10^{-2} - 2.48\times10^{-3}$</td>
</tr>
<tr>
<td>0.7</td>
<td>$-8.40417176$</td>
<td>$-8.38539956$</td>
<td>$-1.88\times10^{-3} - 1.18\times10^{-3}$</td>
</tr>
<tr>
<td>0.8</td>
<td>$-8.5435376$</td>
<td>$-8.5508344$</td>
<td>$-7.97\times10^{-4} - 5.00$</td>
</tr>
<tr>
<td>0.9</td>
<td>$-8.6293186$</td>
<td>$-8.61985420$</td>
<td>$-3.08\times10^{-3} - 1.93$</td>
</tr>
<tr>
<td>1.0</td>
<td>$-8.6675365$</td>
<td>$-8.6659109$</td>
<td>$-1.06\times10^{-3} - 6.67\times10^{-1}$</td>
</tr>
<tr>
<td>1.1</td>
<td>$-8.69263793$</td>
<td>$-8.69231734$</td>
<td>$-3.21\times10^{-4} - 2.01\times10^{-1}$</td>
</tr>
<tr>
<td>1.2</td>
<td>$-8.70632055$</td>
<td>$-8.70624679$</td>
<td>$-7.38\times10^{-5} - 4.63\times10^{-2}$</td>
</tr>
<tr>
<td>1.3</td>
<td>$-8.71367755$</td>
<td>$-8.71367362$</td>
<td>$-3.93\times10^{-6} - 2.47\times10^{-3}$</td>
</tr>
<tr>
<td>1.4</td>
<td>$-8.71756892$</td>
<td>$-8.71757301$</td>
<td>$-4.08\times10^{-6} - 2.57\times10^{-3}$</td>
</tr>
<tr>
<td>1.5</td>
<td>$-8.71954201$</td>
<td>$-8.71954621$</td>
<td>$-4.20\times10^{-6} - 2.64\times10^{-3}$</td>
</tr>
<tr>
<td>1.6</td>
<td>$-8.72047393$</td>
<td>$-8.72047931$</td>
<td>$-2.16\times10^{-8} - 1.36\times10^{-5}$</td>
</tr>
<tr>
<td>1.7</td>
<td>$-8.72088369$</td>
<td>$-8.72088352$</td>
<td>$-1.69\times10^{-7} - 1.06\times10^{-4}$</td>
</tr>
<tr>
<td>1.8</td>
<td>$-8.72102575$</td>
<td>$-8.72102503$</td>
<td>$-7.21\times10^{-7} - 4.52\times10^{-4}$</td>
</tr>
<tr>
<td>1.9</td>
<td>$-8.72107791$</td>
<td>$-8.72107766$</td>
<td>$-2.53\times10^{-7} - 1.59\times10^{-4}$</td>
</tr>
<tr>
<td>2.0</td>
<td>$-8.72107412$</td>
<td>$-8.72107486$</td>
<td>$7.48\times10^{-8} - 4.69\times10^{-4}$</td>
</tr>
</tbody>
</table>

The most important measure is the quality of the projector, as given by the square of the Frobenius norm of the difference between $P_n^2$ and $P_n$, $\|P_n^2 - P_n\|^2$. Each one of these should decrease as the number of iterations, $n$, of the Clinton equations increases.

For the STO-3G basis, the Clinton equations converge to a normalized projector for all separations. Details of each step of the Clinton algorithm for 0.4 $R_{\text{csw}}$ are given in Table 5 along with the energy of the density matrix. The iteration number is denoted by $n$, with $n = 0$ being the energy of the augmented density matrix sum in Eq. (3). As the iteration number, $n$ increases, the parameters measuring the quality of the normalization, idempotency and normalization constraint converge to very small values even in this case, which is the mostclasging. The criterion for convergence of the Clinton algorithm used in this paper was that $\|P^2 - P\|^2 < 1 \times 10^{-28}$. For larger atomic separations in the STO-3G basis, $R > 0.4 R_{\text{csw}}$, the algorithm converges in fewer iterations.

For the proof of concept system in the 6-31G(d, p) basis the Clinton equations converge to a normalized projector in all cases except $R/R_{\text{csw}} = 0.4, 0.5$. In the case for $R/R_{\text{csw}} = 0.4$ the measure of the quality of the projector diverges. The initial value is $\|P^2 - P\|^2 = 2.8$ which increases to $2.4 \times 10^6$ in two steps while the density matrix remain normalized to high precision at each step. Our algorithm abandons further iterations when $\|P^2 - P\|^2 > 1 \times 10^{-6}$. For the case $R/R_{\text{csw}} = 0.5$ the Clinton algorithm gets caught in a local minimum after eight iterations. It remains at $\|P^2 - P\|^2 = 5.95 \times 10^{-5}$ with $E[P] = -2.75$ a.u. for the remaining iterations.

The measures of the quality of convergence for each step of the Clinton algorithm for the 0.6 $R_{\text{csw}}$ 6-31G(d, p) case are given in Table 6. Each column reports the same parameters as in the STO-3G case in Table 5. For larger atomic separations the algorithm converges in fewer steps, as in the STO-3G basis.

As a test of the KEM/KS procedure for a more chemically
interesting system we calculated the KEM energy and apply the KS/KEM procedure to obtain the KS/KEM density for a cluster of twelve water molecules at an energy minimized geometry using KS DFT and the 6-31G(d,p) basis. The results for this case are presented in Tables 7 and 8. The convergence of the Clinton algorithm to a normalized projector is presented in Table 9. The Clinton equations converge rapidly and provide a calculated energy which is closer to the full molecule energy by a factor of nearly four.

4 Conclusions
This is the first study showing how full molecule KS orbitals can be extracted from KS/KEM calculations. The straightforward kernel energy expansion does not deliver full molecule orbitals. To address this gap in the application of the kernel expansion method to solutions of the KS equations we have shown how to construct an initial full molecule KEM density matrix and how to obtain a normalized projector from this initial matrix using the Clinton equations. The resulting matrix, because it is a projector, is factorizable into a product of matrices \( R = C^\dagger C \) whose factors \( C \) will deliver the full molecule KS molecular orbitals \( \phi = C \psi \) expanded in the atomic orbitals. After imposing the projector property upon the density matrix its KS/KEM energy is in the majority of cases closer to that of the full molecule than is the direct KEM energy, as has been shown in the numerical calculations of this paper.

Applied to quantum crystallography \(^{15-17}\), analogously to what has been done here, one may also calculate the KS/DFT quantum mechanics of the kernels composing a molecular crystal. One can then put together the density matrix for entire crystal by properly summing the kernels, at the same time attaching experimental Debye-Waller factors to the basis functions. The resulting density can then be used to calculate the X-ray scattering factors. The magnitude of the resulting crystallographic R-factor then becomes the measure of the crystallographic R-factor then becomes the measure of the accuracy of the KS/KEM orbitals of the molecular crystal. This is important to quantum crystallography in as much as it means that true quantum mechanics, including the exact density KS orbitals can be extracted from the X-ray data.
Acknowledgement: This paper is dedicated to the memory of Professor Robert G. Parr to celebrate his contributions to DFT. We thank Douglas Fox and Fernando R. Clemente of Gaussian, Inc. for invaluable discussions. We thank CUNY for use of computer resources.

References
Appendix D

Submitted Quantum Crystallography publication
Electron Density, Quantum Crystallography, and the Kernel Energy Method*

Walter Polkosnik, (1) Chérif F. Matta, (2-5) Lulu Huang, (6) Lou Massa

(1) Graduate Center, City University of New York, NY 10016, USA. (2) Department of Chemistry and Physics, Mount Saint Vincent University, Halifax, Nova Scotia, Canada B3M 2J6. (3) Dalhousie University, Halifax, Nova Scotia, Canada B3H 1J3. (4) Saint Mary’s University, Halifax, Nova Scotia, Canada B3H 3C3. (5) Département de chimie, Université Laval, Québec, Québec, Canada G1V 0A6. (6) Department of Chemistry Hunter College & the PhD Program of the Graduate Center, City University of New York, NY 10065, USA.

*E-mail: lmassa@hunter.cuny.edu

ABSTRACT

Crystallographic data is the basis for calculation of the complete Quantum Mechanics (QM) of crystallized biological molecules of any size. “Any size” envisions molecules of up to many hundreds of thousands of atoms in a biological complex. The kernel energy method (KEM) has made this doable with any chemical model of any chosen accuracy. This may be of some significance within crystallography made possible by KEM. What we must recognize now is that KEM is capable of providing the entire quantum mechanics of crystallographic biological molecular systems. By this it is meant that given the X-ray structure the KEM formula delivers the energy $E$, and density matrices ($\rho_2$ and $\rho_1$) and the truly quantum mechanical electron density $\rho$. The KEM formula approximations to $E$, $\rho_2$, $\rho_1$, and $\rho$, and to the X-ray structure factors $\{F(K)\}$ are given in this paper. The energy and the density matrices above, of proven KEM accuracy, are sufficient for the complete ground state quantum mechanics of any molecular system consistent with the X-ray structure factors. The algorithm suggested is: (i) Recognize that the KEM formula delivers $E$, $\rho_2$, $\rho_1$, and $\{F(K)\}$; (ii) given a crystal structure, cut it into kernels; (iii) calculate the kernels in the chemical model most appropriate to the accuracy needed; (iv) calculate the full molecule energy $E$ and the density matrices $\rho_2$, $\rho_1$, and $\rho$ using their KEM formulas; (v) from the density matrices calculate any quantum

* Patterned upon a talk by LM at the American Chemical Society’s Middle Atlantic Regional Meeting (MARM) at College of Mount Saint Vincent, Riverdale, NY (June 9-12, 2016).
mechanical property of interest for the whole molecule, including the X-ray structure factors \( \{F(K)\} \) that can be gauged against the experimental ones. The smallness of the crystallographic \( R \)-factor adjudicates the accuracy of the quantum mechanics extracted from the crystallography.

1. INTRODUCTION

Here is a suggestion, obvious but also of some importance. Applied to Quantum Crystallography (QCr),\(^{1-13}\) the kernel energy method (KEM)\(^{10-24}\) implies the use of the electron density to extract the complete quantum mechanics of a crystallized biomolecule. The possibility of such study is much to be desired, because for truly huge molecular systems it would allow the illumination of important biological problems by using the power of true \textit{ab initio} quantum mechanical explanation.

Most crystal structures are brought to final resolution based upon a model equivalent to a sum of spherical atomic electron densities (the independent atom model (IAM)). That molecular density is:

\[
\rho_{\text{IAM}} = \sum_{i=1}^{n} \rho_i^{\text{spherical}}
\]  

(1)

The atoms are positioned so that the crystallographic agreement factor (\( R \)-factor) is minimized:

\[
\delta R = \delta \sum_{K} w_K [F_{\text{calc.}}(K) - F_{\text{obs.}}(K)]^2 = 0
\]  

(2)

where the structure factor is defined as,

\[
F(K) = \int e^{iK \cdot r} \rho(r) d^3r,
\]  

(3)

where \( K \) is the X-ray scattering vector in reciprocal space with components \( \{h, k, l\} \).

Recognizing that bonding will introduce nonspherical aspects to the molecular density, a better density is obtained by a chemical model which incorporates nonspherical density terms. Thus non-spherical multipole representations of the density have been proposed and in use for over four decades.\(^{25-31}\) Hansen and Coppens proposed the now oft used model, which given in standard notation reads:\(^{28,29}\)
\[
\rho_{\text{atom}} = P_{\text{core}}\rho_{\text{core}} + P_{\text{valence}}\kappa^2\rho_{\text{valence}}(\kappa r) + \sum_{l=0}^{l_{\text{max}}} \kappa l^3 R_l(\kappa' r) \sum_{m=-l}^{l} P_{lm} Y_{lm}(\theta, \varphi) ,
\]

where the total number of electrons of the atom or ion \( N \) is broken down into sub-populations \( (P_{\text{core}} + P_{\text{valence}} + \sum_{m=0}^{l_{\text{max}}} P_{lm} = N) \), and where \( \rho_{\text{core}} \) and \( \rho_{\text{valence}} \) are the normalized densities of the free atom or ion, and \( R_l \) are exponential radial functions. The third (valence density) term provides the non-spherical deformation flexibility. This model results in better molecular structures and better electron densities. High quality crystals and the data collection at low temperatures followed by crystallographic refinement within this model yielded high resolution electron densities that account for the topography of chemical bonding.\(^{32-39}\)

An example of a multipole deformation density (defined as the difference between a promolecular density obtained by superposing spherical atomic densities and a multipolar density) is given by that shown in Fig. 1. The figure, obtained from Ref. \([32]\), contains contours of a nonspherical deformation density in a peptide plane. Non-spherical aspects of the density are evident in this image.

The use of the multipolar representation of the density has become a highly regarded (standard) crystallographic technique.\(^{29,30,33}\) What more can be asked in the way of improvements? One possible answer would be to extract the complete quantum mechanics from X-ray scattering data.\(^{12,13}\) Is there a direct connection between X-ray structure scattering data and the quantum mechanics of molecules? Yes, if care is taken to ensure mathematically that the density obtained from crystallography is related to a wavefunction. That relationship is formally called \( N \)-representability, which, symbolically, can be written as:

\[
\rho(1,1') \leftrightarrow \int \Psi^* \Psi d^2d3...dN \gg f(1,1') ,
\]

where the reader is asked to notice the bijective relation between \( \rho \) and \( \Psi \) that does not apply for an arbitrary function \( f \). Stated differently, a density matrix is \( N \)-representable if it can be shown to arise from an antisymmetric \( N \)-body wave function. In particular for:

\[
\rho_1 = tr \mathbf{P}\psi(r)\psi^\dagger(r),
\]
if $P^2 = P$, it may be shown that $\rho_1$ is single determinant \( N \)-representable.\textsuperscript{40} Thus a description for extraction of single determinant quantum mechanics from X-ray scattering may be summarized as follows:

\[
P^2 = P \quad \text{(7)}
\]

\[
tr P = N \quad \text{(8)}
\]

\[
tr Pf(K) = F(K) \quad \text{(9)}
\]

in which case Eq. (6) delivers a density which is both \( N \)-representable and consistent with the experimental scattering data.\textsuperscript{40}

An example of the theoretical quantum crystallography program just described was an application to the Beryllium crystal.\textsuperscript{41} Using simply two basis functions to represent the valence orbital of the Be atom , and the frozen core density from the X-ray tables, the highly accurate data of Larsen and Hansen were used as constraints to fix the single determinant \( N \)-representable density.\textsuperscript{42} The result was one of high accuracy as measured by the remarkably small \( R \)-factor of 0.0018. Furthermore, the errors of the density fall exactly within the statistics expected for a random distribution of errors, taking into account the full range of scattering.\textsuperscript{41}

2. A NEW IDEA FOR PROCEEDING TO BIOLOGICAL MOLECULES

It is seen therefore from the above mentioned beryllium example that true quantum mechanics can indeed be extracted from the X-ray scattering experiment. However, the beryllium “molecule” implies a very small number of electrons. One may ask, what about the case of very large biological molecules? Will exactly the same techniques applied to beryllium serve equally well for thousands or many tens of thousands of atoms, as may occur in biological molecular systems? Can one then, in the same way, find $P^2 = P$ that solves the X-ray refinement problem?

The variation of $P$ against the measured structure factors $F(K)$ becomes impractical as the number of atoms becomes large. For large molecules the number of matrix elements of $P$ increases with the square of the number of atoms, while the number
of X-ray data tend to increase directly with the number of atoms. Clearly there comes a crossover point with increasing numbers of atoms such that insufficient X-ray data are available to deliver unambiguously an $N$-representable density matrix.

A practical way to avoid the dilemma posed by an ever increasing number of atoms is to invoke the Born-Oppenheimer approximation. This asserts that the quantum mechanical electronic structure can be computed at fixed nuclear positions so long as they are known. But the number of X-ray scattering data are sufficient to determine the nuclear positions, so long as the problem is not compounded simultaneously with the many more parameters demanded by a representation of the electron density. Thus it is suggested here to take the nuclear coordinates from the experimental X-ray scattering data and calculating the corresponding electronic structure by the methods of quantum chemistry.43-46

Once the problem of extracting quantum mechanics from the X-ray scattering data is divided into experimental determination of the atomic coordinates and the theoretical calculation of the electronic structure, another problem is confronted. And that is that the molecular quantum mechanical computational burden rises as a high power of the number of basis functions used to expand the molecular orbitals. Obviously for large enough molecules, this burden becomes prohibitive. However, another mathematical solution presents itself in the form of the quantum Kernel Energy Method (KEM), which we now discuss.

3. THE KERNEL ENERGY METHOD (KEM)

The KEM makes possible the calculation of true *ab initio* quantum mechanics of large biological molecules of practically any size with high accuracy.14 This method proceeds by mathematically cutting a large molecule into practicable smaller pieces called “kernels”. As an illustrative example, consider the tripeptide Ala-Phe-Thr. This molecule will be broken into three single kernels each a separate amino acid, so the single kernels are Ala, Phe, and Tyr. In order to account for the pairwise interactions between these kernels, calculations are also performed on double kernels. In this case these will consist
of two chemically contiguous kernels (Ala-Phe) and (Phe-Thr) and one non-contiguous
double kernel (Ala...Thr). The kernels are calculated in the exact geometry as extracted
from that of the full target molecule with appropriate capping by hydrogen atoms to
satisfy the dangling bonds (the contributions of these hydrogens cancel in the KEM
formulae). Since these kernels are calculated as individual molecules, and importantly,
the calculation is inherently parallelizable with possible gigantic saving in computational
efficiency.

The kernels are made by cutting across single bonds in the large molecules, and
hydrogen caps are added to the kernels to preserve the valence of the atoms at the
position of the cuts. As an example for a property that can be predicted from KEM, the
total energy of the above tipeptide is reconstructed from the energy of the kernels
according to the formula:

$$E_{\text{total}} = (E_{\text{Ala}} + E_{\text{Phe}} + E_{\text{Thr}}) + (\Delta E_{\text{Ala-Phe}} + \Delta E_{\text{Phe-Thr}} + \Delta E_{\text{Ala...Thr}}),$$  \hspace{1cm} (10)

where the first summation is of the energies of the single kernels (the separate amino
acids in this case) and the second summation is over all interaction energies of pairs of
amino acid residues. The pair-wise interaction energy, say for the first double kernel
(Ala-Phe), is written,

$$\Delta E_{\text{Ala-Phe}} = E_{\text{Ala-Phe}} - (E_{\text{Ala}} + E_{\text{Phe}}).$$  \hspace{1cm} (11)

When this last equation is substituted into the expression for the full KEM energy one
obtains (in this case):

$$E_{\text{Ala-Phe-Thr}} = (E_{\text{Ala-Phe}} + E_{\text{Phe-Thr}} + E_{\text{Ala...Thr}}) - (3-2)(E_{\text{Ala}} + E_{\text{Phe}} + E_{\text{Thr}}),$$ \hspace{1cm} (12)

and in general, for a molecule broken into \(m\) single kernels, the KEM energy is written:

$$E_{\text{total}} = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} E_{ab} - (m-2) \sum_{c=1}^{m} E_{c}$$ \hspace{1cm} (13)

This is the working equation for the KEM.

The KEM approximation has been proposed as a means to obtain fast approximations
for electron densities and the underlying and related density matrices.12,13,24
\[
\rho_2 = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} \rho_{2(ab)} - (m-2) \sum_{c=1}^{m} \rho_{2(c)}, \quad (14)
\]

\[
\rho_1 = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} \rho_{1(ab)} - (m-2) \sum_{c=1}^{m} \rho_{1(c)}, \quad (15)
\]

\[
\rho = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} \rho_{(ab)} - (m-2) \sum_{c=1}^{m} \rho_{(c)}, \quad (16)
\]

These last three equations represent what may be called the extraction of the complete quantum mechanics from the measured X-ray scattering. The position of the atoms is taken from the crystallography experiment and the electronic structure is calculated by invoking the Born-Oppenheimer approximation followed by the kernel energy method. The quantum mechanics is complete in the sense that all quantum operators of normal interest will have expectation values obtainable from the density and density matrices that have been calculated in the KEM formalism. For example the full molecule electronic energy is of the form

\[
E = \left\langle \hat{T} \rho_1 \right\rangle + \left\langle \hat{V}_{ea} \rho \right\rangle + \left\langle \hat{V}_{ee} \rho_2 \right\rangle \quad (17)
\]

which requires \( \rho_2, \rho_1, \) and \( \rho \) as indicated in the last equation.

Properties of considerable current interest are the total molecular electron density and the total molecular electrostatic potential (ESP) of large macromolecules of biological and nano-technological interest. KEM has been shown to be capable of delivering accurate electron densities sampled at bond and ring critical points (BCPs, and RCPs) in addition to the complete localization-delocalization matrix\(^{47,48}\) of a graphene nanoribbon.\(^{22}\) The method has been extended to estimate electric field-induced changes in the properties (\( i.e., \) response properties) of a finite graphene flake.\(^{23}\) There is, thus, strong evidence supporting the confidence in that the extension of KEM to estimate the total molecular electron density and the total molecular electrostatic potential is only awaiting automation, a current interest in our group.

The KEM approximation to the electron density scalar field of a macromolecule can be numerically achieved from cubes of kernel and double kernel electron densities manipulated point-by-point according to Eq. (16). This KEM density can be inserted in
the expression for the molecular electrostatic potential\textsuperscript{49-54} in which the first term comes from the known molecular structure:

\[ V_{\text{KEM}}(\mathbf{r}) = \sum_{i=1}^{N} \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} - \int \frac{\rho_{\text{KEM}}(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}', \]  
(18)

and where in the first sum the terms with \( \mathbf{R}_i = \mathbf{r} \) are eliminated. The reconstructed electrostatic potential (\( V_{\text{KEM}}(\mathbf{r}) \)) can alternatively be obtained from cubes of ESP of single and double kernels manipulated point-wise using an equation of the same form as Eq. (16).

The KEM has been of tested accuracy over a wide range of biological and other large molecules (such as graphene) and has been shown to be accurate and within a range of commonly used chemical models.\textsuperscript{14} Importantly, Eq. (14-16) imply that the KEM is implementable within any chemical model of choice for the particular problem at hand. It is not inherently restricted to any form of the wavefunction such as a single determinant form.

4. CONCLUSIONS

The principal point of this paper is that a practical way to extract the true quantum mechanics of biological molecules from the X-ray scattering experiment is to invoke the Born-Oppenheimer approximation. In so doing one takes that atomic structure from the X-ray experiment. One recognizes that KEM delivers \( \rho_2, \rho_1, \) and \( \rho \) and subsequently \( F(\mathbf{K}) \), the structure factors. Having the X-ray atomic structure allows creating the kernels mathematically and subsequently calculating their \textit{ab initio} electronic structures. Having the electronic structures of the kernels allows the electronic structure of the full molecule to be reconstructed with KEM accuracy according to the its formulas for \( \rho_2, \rho_1, \) and \( \rho \) [Eqs. (14-16)].

It follows that all expectation values for the full molecule are calculable from these density matrices as shown for the energy, as one example [Eq. (17)]. And finally if the X-ray structure factors \( \{F(\mathbf{K})\} \) are calculated with the KEM density \( \rho(\mathbf{r}) \) [Eq. (3)]
that gives the ultimate confirmation of the accuracy of the quantum mechanics obtained from the experiment.

Fig. 2 summarizes the conceptual connection of KEM within the context of quantum crystallography. The figure is to be read starting from the X-rays source that yields the diffraction data that is used to generate the full molecular geometry. The molecule is then partitioned into kernels, sufficiently small to be subjected to quantum chemical computation – each kernel (or double kernel) separately (and hence highly parallelizable). The calculated density matrices of the kernels are then combined to yield those of the full molecule according to Eqs. (14-16). The KEM density can then be Fourier transformed to deliver calculated structure factors. The calculated and experimental structure factors can be compared by a crystallographic $R$-factor.

ACKNOWLEDGMENTS

Funding for this project was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI), and Mount Saint Vincent University - (C. F. M.), the U.S. Naval Research Laboratory (project # 47203-00 01) and by a PSC CUNY Award (project # 63842-00 41) - (L. M.); and the Office of Naval Research (ONR) through the Naval Research Laboratory's Basic Research Program - (L. H.)
Fig 1. Multipolar deformation electron density in the plane of a peptide bond (N4-C13-O4) obtained by Lecomte et al.\textsuperscript{32} (Left) The experimental density with contour interval of 0.05 eÅ\textsuperscript{3} and where solid lines indicate positive contours and dashed ones indicate negative contours. The nodal contour has been omitted. (Reproduced with permission of the copyright holder. © 1992 American Chemical Society).
Fig 2. A conceptual sketch of the quantum crystallography/kernel energy method (QCr/KEM) scheme from X-rays scattering by the molecule to its fragmentation into pieces passing through the quantum mechanical KEM reconstruction of its density matrices. The reconstructed density and calculated $F(K)$ can then be gauged against experiment. (Reproduced with permission of the copyright holder from Ref. 55 © 2017 Springer).
References


(24) Huang, L.; Matta, C. F.; Massa, L. The kernel energy method (KEM) delivers fast and accurate QTAIM electrostatic charge for atoms in large molecules. Struct. Chem. 2015, 26, 1433-1442.


(31) Macchi, P.; Gillet, J.-M.; Taulelle, F.; Campo, J.; Claiser, N.; Lecomte, C.
Modelling the experimental electron density: only the synergy of various approaches can tackle the new challenges. *Int. U Cryst. J. (IUCrJ)* 2015, **2**(Pt 4), 441–451.


(49) Bonaccorsi, R.; Scrocco, E.; Tomasi, J. Molecular SCF calculations for the ground state of some three-membered ring molecules: (CH$_2$)$_3$, (CH$_2$)$_2$NH, (CH$_2$)$_2$NH$_2^+$, (CH$_2$)$_2$O, (CH$_2$)$_2$S, (CH)$_2$CH$_2$, and N$_2$CH$_2$. *J. Chem. Phys.* 1970, **52**, 5270-5284.


(54) Tomasi J.; Cappelli, C.; Mennucci, B.; Cammi, R. *Quantum Biochemistry: Electronic Structure and Biological Activity (Volume 1)*; Wiley-VCH: Weinheim, 2010; Chapter <24 Issue

Bibliography


[38] L. Huang, C. Matta, and L. Massa, Structural Chemistry 26, 1433 (2015).


