Symposium: New Generation Quantum Theory - Particle Physics, Cosmology, and Chemistry -
University of Kyoto, March 7-9, 2016

Invited Lecture

Dedicated to the memory of Professor Kenichi Fukui

Relations Between the Part and the Whole in Quantum Chemistry

Paul G. Mezey

Canada Research Chair in Scientific Modeling and Simulation
Department of Chemistry and Department of Physics and Physical Oceanography
Memorial University of Newfoundland
283 Prince Philip Drive, St. John's, NL, CANADA A1B 3X7
Tel.: 1 709 737 8768, Fax: 1 709 737 3702,
email: paul.mezey@gmail.com
http://www.mun.ca/research/chairs/mezey.php
明日試験

井謙一博士
その人と学问

平成10年4月10日（土）
〜5月23日（土）

京都大学総合博物館
Professor Kenichi Fukui, a scientist, and engineer, and a humanist

He was an exceptional man, a man of science, and a man of culture.

I could never forget the spiritual effect I felt, when on an excursion to Mount Hiei with his research group, I could walk a few steps with a walking stick he had received as a gift from a Buddhist Monk, who himself had made the thousand night mountain climbing pilgrimage to Mount Hiei.

He was a thorough theorist, but also an engineer, who discovered not only fundamental theories, but also the practical aspects of new theories.

Just two examples:

The IRC, the Intrinsic Reaction Coordinate, developed together with Professor Akitomo Tachibana

The Frontier Orbital Theory and Method,

Both of these theories and practical methods are used by hundreds of scientists
Relations Between the Part and the Whole in Quantum Chemistry

Molecular information, fragment information, holographic relation

Fuzzy molecular fragment studies for local effects
Substituent effects and local shape
Through-bond and through-space interactions
Functional groups, their shape variations linked to effect variations
Fragment-based macromolecular quantum chemistry computations
  efficient linear-scaling methods for proteins
Combinatorial quantum chemistry
  efficient tools for quantum chemical molecule design
ABSTRACT

According to the **Holographic Electron Density Theorem**, “Any nonzero volume *part* of a molecular electron density in a non-degenerate ground state contains the *complete* information about all properties of the *entire, boundaryless molecule*” [1].

One extension of this theorem, the **Holographic Electron Density Theorem for Latent Properties** states: “Any nonzero volume *part* of a molecular electron density in a non-degenerate ground state contains the *complete* information about *all latent, non-exhibited properties* of the entire, boundaryless molecule” [2].

The limitations and some useful shortcuts of fragment-based macromolecular quantum chemistry are based on these theorems, interrelating the part and the whole of molecular systems.

Some of the computational methodologies for **macromolecular “fuzzy” fragment density matrix methods**, the ADMA and the **macromolecular conformation analysis method**, LIL-ADMA will be discussed in connection to these theorems [3-9].


Natural molecular fragments, functional groups, and holographic constraints on electron densities†

Paul G. Mezey*

Received 23rd January 2012, Accepted 27th April 2012
DOI: 10.1039/c2cp40237h

One of the tools of the shape analysis of molecular electron densities, the Density Threshold Progression Approach used in Shape Group studies can also serve as a criterion for the selection of “natural” molecular fragments, relevant to functional group comparisons, reactivity studies, as well as to the study of levels of relative “autonomy” of various molecular regions. The relevance of these approaches to the fragment-based studies of large molecules, such as biopolymers and nanostructures is emphasized, and the constraints represented by the holographic electron density theorem to this and alternative recent fragment approaches are discussed. The analogies with potential energy hypersurface analysis using the Energy Threshold Progression Approach and connections to level set methods are discussed, and the common features of these seemingly distant problems are described.
Fuzzy Electron Density Fragments in Macromolecular Quantum Chemistry, Combinatorial Quantum Chemistry, Functional Group Analysis, and Shape–Activity Relations

Paul G. Mezey*

Scientific Modelling and Simulation Laboratory, Department of Chemistry, Memorial University of Newfoundland, St. John’s, Newfoundland A1B 3X7, Canada

CONSPECTUS: Just as complete molecules have no boundaries and have "fuzzy" electron density clouds approaching zero density exponentially at large distances from the nearest nucleus, a physically justified choice for electron density fragments exhibits similar behavior. Whereas fuzzy electron densities, just as any fuzzy object, such as a thicker cloud on a foggy day, do not lend themselves to easy visualization, one may partially overcome this by using isocontours. Whereas a faithful representation of the complete fuzzy density would need infinitely many such isocontours, nevertheless, by choosing a selected few, one can still obtain a limited pictorial representation. Clearly, such images are of limited value, and one better relies on more complete mathematical representations, using, for example, density matrices of fuzzy fragment densities. A fuzzy density fragmentation can be obtained in an exactly additive way, using the output from any of the common quantum chemical computational techniques, such as Hartree–Fock, MP2, and various density functional approaches.

Such "fuzzy" electron density fragments properly represented have proven to be useful in a rather wide range of applications, for example, (a) using them as additive building blocks leading to efficient linear scaling macromolecular quantum chemistry computational techniques, (b) the study of quantum chemical functional groups, (c) using approximate fuzzy fragment information as allowed by the holographic electron density theorem, (d) the study of correlations between local shape and activity, including through-bond and through-space components of interactions between parts of molecules and relations between local molecular shape and substituent effects, (e) using them as tools of density matrix extrapolation in conformational changes, (f) physically valid averaging and statistical distribution of several local electron densities of common stoichiometry, useful in electron density databank mining, for example, in medicinal drug design, and (g) tools for combinatorial quantum chemistry approaches using fuzzy fragment databanks and rapid construction of a large number of approximate electron densities for large sets of related molecules, relevant in theoretical molecular and nanostructure design.
Relations between real molecules through abstract molecules: the reference cluster approach

Paul G. Mezey\textsuperscript{1,2,3}

Received: 15 June 2015 / Accepted: 13 September 2015 / Published online: 16 October 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Replacing integer atomic numbers representing nuclear charges by continuous variables has already provided some rigorous quantum chemical relations between real molecules, using a formal interpolation through non-physical abstract molecules of continuously varying nuclear charges. Extending this approach to the more general “universal molecule” model, where all discrete parameters of molecules are generalized and replaced by continuous variables, provides further relations, actually interconnecting all real molecules through abstract, non-physical “molecules,” involving, for example, non-integer number of electrons. One simplifying idea of this model is the so-called “reference cluster,” originally defined for isoelectronic systems of a fixed number \( N \) of nuclei, where each nucleus is replaced by a possibly fictive nucleus with a nuclear charge equal to the average of the \( N \) nuclear charges. Based on the earlier results, some new relations are derived interrelating energies and some other properties of some real molecules, and also providing a unifying framework for the utilization of both symmetry and energy relations of the universal molecule model.

Keywords Nuclear charge space · Electronic energy inequalities · Reference cluster · Universal molecule model

1 Introduction

Originally motivated by the United Atom studies of Thirring, Narnhofer, and Lieb [1–3], the idea of exploiting nuclear charge convexity relations between electronic energies of molecules, without involving the extreme case of uniting all nuclei of the molecule into a single, formal nucleus of the United Atom, has lead to a variety of rigorous quantum chemical energy inequalities for various sets of molecules, starting with simple diatomic cases [4, 5].

After the very first examples of such electronic energy inequalities [4, 5], including one of the simplest cases, the \( \text{N}_2 \) and \( \text{CO} \) pair, the concept of the nuclear charge space was elaborated [6] and the level set topology approach was introduced [7]. A generalization to other linear parameters besides the nuclear charges was described [8], followed by several additional applications [9–15]. As one important
Electron density shape analysis of a family of through-space and through-bond interactions

Zoltan Antal, Peter L. Warburton and Paul G. Mezey*

A family of styrene derivatives has been used to study the effects of through-space and through-bond interactions on the local and global shapes of electron densities of complete molecules and a set of substituents on their central rings. Shape analysis methods which have been used extensively in the past for the study of molecular property–molecular shape correlations have shown that in these molecules a complementary role is played by the through-space and through-bond interactions. For each specific example, the dominance of either one of the two interactions can be identified and interpreted in terms of local shapes and the typical reactivities of the various substituents. Three levels of quantum chemical computational methods have been applied for these structures, including the B3LYP/cc-pVTZ level of density functional methodology, and the essential conclusions are the same for all three levels. The general approach is suggested as a tool for the identification of specific interaction types which are able to modify molecular electron densities. By separately influencing the through-space and through-bond components using polar groups and groups capable of conjugation, some fine-tuning of the overall effects becomes possible. The method described may contribute to an improved understanding and control of molecular properties involving complex interactions with a possible role in the emerging field of molecular design.
Substituent effects and local molecular shape correlations

Zoltan Antal and Paul G. Mezey*

Using a detailed electron density shape analysis methodology, a new method is proposed for studying the main components of substituent effects in a series of disubstituted benzenes, in correlation with their activating and deactivating characteristics as observed by the induced shape changes of a local electron density cloud. The numerical measures obtained for the extent of shape changes can be correlated with known and with some unexpected effects of various substituents. The insight obtained from the shape analysis provides a theoretical, electron density based justification for some well-known trends, but it also provides new explanations for some of the unexpected features of these substituent effects.
First just one practical motivation for studying molecular fragments: Through-bond and through-space interactions within molecules.

and, e.g., para-metoxy-styrene
Some practical motivation for the study of molecular fragments: Through-bond and through-space interactions within molecules.

In most cases, it is hard to study these interactions separately, but in some special cases, it is possible.

The case of para-substituted styrene molecules, two approaches:

Study the fragment shape changes for the vinyl group, as influenced by the para substituent

(a) In the complete molecule
(b) In a pair of molecules where the benzene ring is “left out”

Case (a) includes both through-bond and through-space interactions
Case (b) includes only through-space interactions.
Styrene 0.1 a.u.
isodensity contours
Styrene 0.01 a.u.
isodensity contours
Para-metaoxy-styrene  0.1 a.u.
isodensity contours
Para-metoxy-styrene  0.01 a.u.
isodensity contours
The current focus is on four topics:

1. **Holographic properties of electron densities**, relations between the part and the whole in molecules *(Theory)*
2. **Additive Fuzzy Density Fragment (AFDF)** methods exploiting relations between the part and the whole, **Adjustable Density Matrix Assembler (ADMA)** macromolecular methods *(Computation)*
3. **Macromolecular Density Matrix Extrapolation (DME)**, and the **Löwdin - Inverse - Löwdin (LIL)** transformation method for macromolecular conformation analysis *(Computation)*
4. **The Universal Molecule Model** for interrelations among various molecules *(Theory and Computation)*
Brief summary of the
Holographic Electron Density Theorem
1. What is the connection to Holography? Only analogy:

The Part contains all information about the Whole.

Optical holographic recording:
the complete 3D image can be recovered
from any non-zero area piece of the holographic plate

Holographic Properties of Electron Densities

and

Holographic Principles on Potential Energy Surfaces:

Ideal tools for quantum chemical extrapolations!
Holographic Electron Density Theorem, (a strengthening of the Hohenberg-Kohn theorem):

For a non-degenerate ground state, the complete molecular information is encoded in any small positive volume part of the electron density cloud. Each part “knows” everything about the whole!


One extension for **Latent Properties:**


For a ground state molecule, a property of one of its excited states is a **latent property**.

For conformation A, a property of another conformation B is one of the latent properties of A!

A tool for potential surface extrapolation!
A paper on a universal relation between the part and the whole:


The holographic electron density theorem and quantum similarity measures

PAUL G. MEZÉY

Institute for Advanced Study, Collegium Budapest, 1014 Budapest, Szentháromság u. 2, Hungary, and Mathematical Chemistry Research Unit, Department of Chemistry and Department of Mathematics and Statistics, University of Saskatchewan, 110 Science Place, Saskatoon, SK, Canada S7N 5C9

(Received 26 January 1998; accepted 12 March 1998)

How much information about the complete molecule is present in a part of the molecule? Quantum similarity measures provide comparisons between molecular electron densities based on integration over the whole space. Such integration involves boundaryless electron densities, whereas an early application of the Hohenberg-Kohn theorem to local subsystems of molecules requires these molecules to be confined to bounded, finite regions of the space. However, actual molecules have no boundaries, they are not confined to any finite region of the space. In order to find deterministic relations between local and global, boundaryless electron densities, and to classify the link between quantum similarity measures involving the full space and local subsystems, the unique extension property called the holographic property of subsystems of complete, boundaryless electron densities is established. Any nonzero volume piece of the ground state electron density completely determines the electron density of the complete, boundaryless molecule.
Electron density and the information content of molecules

The Hohenberg-Kohn Theorem, 1964: *The molecular electron density determines the molecular energy and through the Hamiltonian, all other molecular properties.*

That is, the electron density is the fundamental information storage of molecules, the electron density actually carries the complete information about the molecule.


The holographic information content in *parts* of molecules:

“The Holographic Electron Density Theorem”, Paul G. Mezey 1999:

*Any nonzero volume part of a molecular electron density in a non-degenerate ground state contains the complete information about all properties of the entire, boundaryless molecule.*


This fundamental property of all molecules applies to all exhibited and also to all latent molecular properties.
Girona, 2005, Walter Kohn and Paul G. Mezey

Holographic Electron Density Theorem, HEDT, Paul G. Mezey
Historical notes on the Holographic Electron Density Theorem

The Hohenberg-Kohn Theorem [1] refers to the complete electron density, stating that all non-degenerate, ground state molecular properties are determined by the complete electron density.

The “Holographic Electron Density Theorem”[2] deduces complete information from the part, stating that any nonzero volume part of a molecular electron density in a non-degenerate ground state contains the complete information about all properties of the entire, boundaryless molecule.

That is, the complete molecular information does not require the complete electron density, and local electron density ranges already fully determine all molecular properties.

For artificial, bounded systems, that is, not for molecules, an earlier result on the relation between the part and the whole has been proven [3], by J. Riess and W. Münch. Their proof was not applicable for real, boundaryless molecules. This limitation of the result, although clearly stated by these authors, has not always been fully recognized by some later papers quoting the result.

Is the distinction between

(1) finite, closed molecular models with boundaries, and

(2) open, boundaryless molecular models

really so important?

YES !
Two deadly sins:

1. Division by zero

2. Modeling open systems as if they were closed systems

Doing either, you may “prove” anything and the opposite, simultaneously

(thereby achieving the ultimate democracy of ideas).

What is the Ultimate Democracy?
“All Numbers Are Created Equal”.

(P.G. Mezey)

(The above remark was the source of a witty poem by Professor K. Balasubramanian)
Some powerful theorems, such as some of those on analytic continuation, do not apply directly to open, boundaryless, infinite systems, unless some special transformations are applied.

One such transformation is the Alexandrov one-point compactification.

(Compactness is a generalization of the property of “closed and finite “).
Figure 1. Three-dimensional illustration of the Alexandrov one-point compactification method, as applied to the two-dimensional plane $E^2$, resulting in the sphere $S^2$. Point $p$ of distance $r$ from the origin $o$ is assigned to the unique point $p'$ on the sphere $S^2$ using the straight line issued from the north pole $n$ of sphere $S^2$ to point $p$. The polar angle $\theta$ of point $p'$ is also indicated. A four-dimensional version of this method leads to a complete molecular electron density function that is analytical almost everywhere on a sphere $S^3$.
The actual proof of the Holographic Electron Density Theorem avoids all deadly sins, it involves some topological tools and is based on circumventing the problem of non-boundedness of molecules, by using

(i) a four-dimensional representation (fourth dimension is the electron density) and a one-point compactification of the 3D space to a 3D sphere embedded in 4D, and

(ii) the proof that on this sphere all points of analyticity of the density function are connected, and

(iii) the proof that the points of non-analyticity form a measure zero set.

Various consequences of the holographic theorem provide

(i) symmetry constraints,
(ii) new tools for Molecular Informatics and
(iii) for the prediction of latent molecular properties, such as those of excited electronic states and in alternative conformations [1],
(iv) new extensions of molecular shape analysis[2]
(v) and the linear scaling, \textit{ab initio} quality macromolecular quantum chemistry computations[3,4] based on the additive fuzzy density fragmentation (AFDF) approach[3,4],
(vi) with applications in macromolecular shape - activity correlations involving large biochemical systems.


In the direct structure solution of noncentrosymmetric crystals, a partial structure rather than a complete structure is obtained from the initial phase determination. Using the phases generated by the partial structure as input into Karle’s tangent formula, provided a procedure for generating the complete structure, when the fragment had been located in the unit cell.
Local density contours influenced by the rest of the molecule:

Four fuzzy electron density fragments of phenanthrene shown at density thresholds 0.001au, 0.01au, and 0.1au. The location of each carbon atom is given in parentheses.

- C(4a)
- C(8a)
- CH(2)
- CH(9)
The main tool:

the Holographic Electron Density Theorem can be applied to
Latent Molecular Properties,

especially, to extrapolation on potential surfaces
In a nutshell (well, in a coconut shell):

conformers do not keep secrets from each other,
they are exceptionally good neighbors,
they know everything about one another.

It is enough if you ask only one of them,
this conformer will tell you everything about all its neighbors.
LATENT MOLECULAR PROPERTIES

Take a molecule $A$ in its electronic ground state.

Typically, the molecular properties exhibited by the isolated molecule $A$ are those which are primarily associated with nuclear arrangements similar to the most stable one (energy minimum).

For molecule $A$, properties associated
• with electronically excited states, or
• with highly distorted nuclear arrangements very different from the most stable one, or
• with other stable or unstable molecules of the same overall stoichiometry, etc.

are regarded as LATENT PROPERTIES, exhibited only in special instances.
Take a molecule $A$ in its (non-degenerate) electronic ground state. We know from the Holographic Electron Density Theorem that the complete information on all actual properties of molecule $A$ is contained within any small, nonzero volume piece of the electron density cloud of molecule $A$.

According to the Holographic Principle for Latent Molecular Properties, much more is true:

**Any small, nonzero volume piece of the electron density cloud of molecule $A$ contains the complete information on all actual and latent molecular properties of molecule $A$.**

Latent Property Prediction

For example, any small, nonzero volume piece of the non-degenerate ground state electron density cloud of molecule A contains all information on

- all electronically excited states of molecule A
- any, even highly distorted nuclear arrangements very different from the most stable one,
- any other stable or unstable molecules of the same overall stoichiometry.

In fact, this implies that the information on the entire potential energy hypersurface (PES) of any electronic state is contained in any small, nonzero volume piece of the non-degenerate ground state electron density cloud of molecule A.

Any small, nonzero volume piece of the non-degenerate ground state electron density cloud of molecule A “knows” everything what this molecule could possibly do, including all stoichiometry-preserving reactions.
Extrapolation constraint from a single point of the potential energy hypersurface

Extrapolation on potential energy hypersurfaces:

According to the latent property version of the holographic electron density theorem, the (nearly) complete information on the entire potential energy hypersurface (PES) of any electronic state is contained in any small, nonzero volume piece of the non-degenerate ground state electron density cloud of molecule A of a given nuclear configuration K.
Some Consequences of the Holographic Electron Density Theorem on the quantum mechanical relations between the parts of molecules and whole molecules:

Molecular Fragment Analysis,

New Approaches to Macromolecular Quantum Chemical Calculations
Local Fuzzy Density Fragments provide an alternative to Orbital Localization approaches, such as the Pipek-Mezey approach:
2. From Principle to Practice

(Constraints and utilization)
Brief summary of the macromolecular

ADMA method:

Adjustable Density Matrix Assembler

based on

Additive Fuzzy Density Fragments
An excellent recent review of these and related methods has been given by

Alexey V. Akimov and Oleg V. Prezhdo*


DOI: 10.1021/cr500524c
Advantages of Additive Fuzzy Density Fragments (AFDF):

- No artificial boundaries, fragments mimic fuzzy densities of real molecules
- Applicable for local shape analysis of molecular parts
- Provide tools for macromolecular quantum chemistry

A linear scaling macromolecular method based on AFDF:

- Adjustable Density Matrix Assembler (ADMA)
From Local Information to Global Representation:
Additive Fuzzy Density Fragment (AFDF) Approaches

A Fragment Density Matrix Approach to Linear Scaling
Macromolecular Quantum Chemistry:

The MEDLA and ADMA Methods
MOLECULAR FRAGMENTS FOR LOCAL SHAPE ANALYSIS

QUASI-TRANSFERABLE MOLECULAR FRAGMENTS FOR LINEAR-SCALING

MACROMOLECULAR QUANTUM CHEMISTRY METHODS
Not all fragmentation schemes are advantageous.

Note: electron density fragments with boundaries,

e.g. those obtained by zero flux surfaces can never match, not even by accident, if placed into a different molecule,

the local error there is 100%

(either zero density or double density at forced joints,
at the most crucial bond critical point region).

For fuzzy density fragments of ADMA

the error is never accumulated anywhere,
it can be fully controlled,
one can make it smaller than any threshold by increasing the ADMA distance criterion.
Not all fragmentation schemes are advantageous:
A careless fragmentation scheme may lead to an undesirable
Additive Fuzzy Electron Density Fragmentation, AFDF
Fuzzy electron density decomposition of molecule AB
Fuzzy electron density decomposition of molecule CD

ADMA, Adjustable Density Matrix Assembler, Paul G. Mezey
Fuzzy electron density construction of molecule AD
An alternative based on electron density “pieces” with sharp boundaries: I do not recommend it.
Unavoidable discontinuity of fragments with boundaries when trying to build new molecules

FRAGMENTS DON'T MATCH, DOMAINS OF 100% ERROR

DENSITY BUILDING USING "ATOMS IN MOLECULES" APPROACH:
STEP 1: FRAGMENTATION BY BADER'S ZERO FLUX SURFACES
STEP 2: JOINING FRAGMENTS TO FORM NEW MOLECULES

FRAGMENTATION:

ZERO FLUX SURFACES

A

B

C

D

FRAGMENTS:

A

B

C

D

JOINT GENERATION:

DENSITY "DOUBLING", 100% ERROR

ADMA, Adjustable Density Matrix Assembler, Paul G. Mezey
Problems with molecular fragments based on zero flux surfaces

Some words of caution concerning the AIM approach: it has many well-documented failures, even in small molecules, such as “Bader-atoms” without a nucleus. Some references:

Luiz A. Terrabuio, Tiago Q. Teodoro, Marina G. Rachid, and Roberto L. A. Haiduke*

Systematic Theoretical Study of Non-nuclear Electron Density Maxima in Some Diatomic Molecules

J. Phys. Chem. A 2013, 117, 10489−10496
Avoid models with boundaries. In a fuzzy, additive, but non-exact electron density construction of molecule AD, there is no accumulation of error.
From Local Information to Global Representation: Additive Fuzzy Density Fragment (AFDF) Approaches

A Fragment Density Matrix Approach to Linear Scaling

Macromolecular Quantum Chemistry:

The MEDLA and ADMA Methods
TWO ASPECTS:

Extension of design methods to large structures:

linear scaling methods

Development of tools for analyzing local and global shapes for design purposes:

electron density shape analysis
Extension of design methods to large structures:

linear scaling methods
For the design of large structures, macromolecular quantum chemistry methods compatible with methods applied to smaller systems are needed.

One approach with intuitively clear connection to conventional methods, using a single, transparent distance condition for accuracy:

A Fragment Density Matrix Approach to Linear-Scaling Macromolecular Quantum Chemistry:

The MEDLA and ADMA Methods
From Local Information to Global Representation:

Additive Fuzzy Density Fragment (AFDF) Approaches

A Fragment Density Matrix Approach to Linear-Scaling

Macromolecular Quantum Chemistry:

The MEDLA and ADMA Methods
The pioneer of density partitioning (he called it population analysis), R. Mulliken, near the crater of Teide, Spain, 1976
The Additive Fuzzy Density Fragmentation (AFDF) Principle.

Based on a scheme analogous to Mulliken’s Population Analysis

*without integration.*


[MEDLA reference]


[ADMA reference]
Two “Additive Fuzzy Density Fragment” (AFDF) Macromolecular Linear Scaling Methods

Standard \textit{ab initio} quantum chemical calculations, like the traditional Hartree-Fock (HF) method, are only feasible for small to medium size molecules due to the high power scaling behavior ($O(N^3)$ or worse) with system size.

Two of “Fuzzy Fragment” macromolecular linear scaling methods introduced to circumvent this problem are the MEDLA (Molecular Electron Density Loge Approach, \textbf{1993}, Walker, Mezey) and the more advanced ADMA method (Adjustable Density Matrix Assembler, \textbf{1995}, Mezey), \textit{exactly linear-scaling}, both based on \textit{fuzzy electron density fragment construction and additive assembly}.
Natural vs. technical aspects of linear scaling quantum chemistry

Preference is given to linear scaling approaches based on the natural, fundamental properties of molecules, such as fuzzy, gradually decaying electron density contributions, as opposed to technical, computational concepts such as integral estimates, eigenvalue problem simplifications, and the like.

Distance is the most universal condition nature uses in deciding which interactions are more and which are less important. Use natural distance criteria instead of magnitude estimates for various integrals of various computational schemes.

In order to address the actual physical problem, we may consider

(i) The fundamental traditional chemical concept of Functional Groups

(ii) The quantum chemical relations between the Whole and the Parts of a molecule
Density matrices can be computed using standard Hartree-Fock ab initio methods, and their information content is equivalent that of molecular wavefunctions.

The electron density \( \rho(\vec{r}) \) of a molecule can be expressed in terms of a basis set of \( n \) atomic orbitals \( \varphi_i(\vec{r}) \) (\( i=1,2,...,n \)) used for the expansion of the molecular wavefunction and the density matrix \( P \) of elements \( P_{ij} \) determined for the given nuclear configuration using the specified basis set:

\[
\rho(\vec{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} \cdot \varphi_i(\vec{r}) \cdot \varphi_j(\vec{r})
\]  

(1)
“Fuzzy Fragment” linear scaling Adjustable Density Matrix Assembler (ADMA) method (first three steps):

1. subdivide nuclei of macromolecule into groups,
2. for each group generate a small “parent molecule” containing the given set of nuclei, and also including some surrounding neighbor groups and adding peripheral H (or other) atoms to avoid “dangling bonds”.
3. carry out some standard QM calculation for each parent molecule

Next, do density matrix fragmentation for each parent molecule as follows …..
The Additive Fuzzy Density Fragmentation (AFDF) Principle.

Based on a scheme analogous to Mulliken’s Population Analysis without integration.


1. Subdivide the molecule into a set of \( m \) mutually exclusive families of nuclei denoted by \( f^k, k = 1, \ldots, m \)

2. The fragment density matrix of the nuclear family \( f^k \) is then defined according to the Mulliken-Mezey scheme as:

\[
P^k_{ij} = \begin{cases} 
P_{ij} & \text{if } \varphi_i(\vec{r}) \text{ and } \varphi_j(\vec{r}) \text{ are centered on nucleus of } f^k \\ 
0.5 \cdot P_{ij} & \text{if } \varphi_i(\vec{r}) \text{ or } \varphi_j(\vec{r}) \text{ is centered on nucleus of } f^k \\ 
0 & \text{otherwise} 
\end{cases}
\]

3. By exact additivity, the total density matrix of the molecule is expressed as:

\[
P_{ij} = \sum_{k=1}^{m} P^k_{ij}
\]
Fuzzy Fragment Density Matrix Methods (cont.)

For the $k$-th fragment, the local, fuzzy electron density is defined as

$$\rho^k(\vec{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}^k \cdot \varphi_i(\vec{r}) \cdot \varphi_j(\vec{r})$$

As follows from the exact additivity of the fragment density matrices, these fragment densities are also exactly additive:

$$\rho(\vec{r}) = \sum_{i=1}^{m} \rho^i(\vec{r})$$
“Fuzzy Fragment” linear scaling Adjustable Density Matrix Assembler (ADMA) method (continued):

1. subdivide nuclei of macromolecule into groups,
2. for each group generate a small “parent molecule” containing the given set of nuclei, and also including some surrounding neighbor groups and adding peripheral H (or other) atoms to avoid “dangling bonds”.
3. carry out some standard QM calculation for each parent molecule

After steps 1-3 are done,

4. do fuzzy density decompositions, keep fragment density matrices of the central nuclear families, containing “half” the interactions with each surrounding fragment in the parent molecule, other “half” is obtained when the neighbor fragment is central in a different “parent molecule”.
5. Combine fragment density matrices by simple addition to form approximate macromolecular density matrix.

Method is exactly additive, if applied to a parent molecule, and gives good to excellent approximation (with all short range interactions included) for larger molecules. Increasing parent molecule size can achieve any desired accuracy, within the given Q Chem method.
The Adjustable Density Matrix Assembler (ADMA) Method

1. The nuclei of the ‘target’ macromolecule are subdivided into mutually exclusive families of nuclei denoted by $f^k, k = 1, ..., m$.

2. Generate a set of small parent molecules where the center of each of these parent molecules contains one of the nuclear families $f^k$ with the same local nuclear geometry and the same local surroundings as in the target macromolecule. Each parent molecule also contains the immediately surrounding nuclear families within some distance. The accuracy can be fully controlled by how much of the surroundings is included, that is, by the size of these parent molecules.

3. Carry out conventional (e.g. Hartree-Fock) QM computations for the parent molecules. Note that, for a constant size limit for the parent molecules, the computation time is linear in the number of parent molecules, that is, linear in the macromolecular size.

4. Apply the AFDF procedure for each of the parent molecules in order to generate the fragment density matrices for each central nuclear family.

5. Apply the additivity principle for all fragment density matrices for the complete set of nuclear families in order to generate the macromolecular density matrix.

COMMENT: each computational step is linear in the size of the macromolecule.
Example for target “macromolecule” and two parent molecules

In the upper part of the figure, the target molecule subdivided into 27 fragments is shown with the atoms color-coded according to the assignment to the fragments. In the lower part, two parent molecules for two of these fragments and their 4 Å surroundings are shown. On the left hand side, the central fragment is a CO₂-group and on the right, it is a CH-group, both shown in yellow. In the parent molecules, the atoms of the surroundings are color-coded by atom type. Atoms not linked covalently to the nuclei of the central fragment within the parent molecule, but less than 4 Å away from these central nuclei, are also included in the surroundings. Some hydrogen atoms filling the missing valences are also included.
Additional technical comments on the ADMA method

For any given macromolecule, an automated procedure is used for the generation of the parent molecules and the calculation of the fragment density matrices.

The resulting fragments plus surroundings are saved in pdb-format and converted to a z-matrix file, which is suitable for a Hartree-Fock self-consistent field calculation using the Gaussian 98 or some similar program.

The quantum chemical calculations for all fragments are then initiated automatically.

Note that from each parent molecule only the fragment density matrix corresponding to the central fragment is used in further calculations, hence it is accurate within the limits set by the distance criterion chosen for the surroundings within the parent molecule.

Consequently, the density matrix of the target macromolecule can be approximated to any desired accuracy within the QM method used, if fragment density matrices are taken from large enough parent molecules,

*hence a single size parameter controls the accuracy* of the ADMA method.
ADMA Energy calculations

Within the ADMA approach, the total energy of the target molecule is calculated following the standard Hartree-Fock formalism using the ADMA density matrix replacing the ideal (and for most large molecules still unattainable) exact density matrix of the macromolecule:

\[ E_{HF} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( F_{ij} + H_{ij}^{core} \right) \cdot \sum_{k=1}^{m} P_{ij}^k + V_{NN} \]

\[ F_{ij} = H_{ij}^{core} + \sum_{r=1}^{b} \sum_{s=1}^{b} \left( [ij| rs] - \frac{1}{2} [is| rj] \right) \cdot \sum_{k=1}^{m} P_{ij}^k \]

It has been shown that by using large surroundings of more than 10-12 Å radius are sufficient to achieve truly high quality results, which differ from the results of the standard Hartree-Fock method by less than 1 kcal/mol (0.00156 Hartree).
Equations for calculating various properties using the ADMA method

Table 1. Equations for the Calculations of Approximate Properties Using the ADMA Approach.

Electron Density
\[ \rho(\vec{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi_i(\vec{r}) \cdot \varphi_j(\vec{r}) \cdot \sum_{k=1}^{m} P_{ij}^k \]

Hartree–Fock energy
\[ E_{HF} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( (F_{ij} + H_{ij}^{corr}) \cdot \sum_{k=1}^{m} P_{ij}^k \right) + V_{NN} \text{ with } F_{ij} = H_{ij}^{corr} + \sum_{r=1}^{b} \sum_{s=1}^{b} \left[ (\langle \psi | s \rangle - \frac{1}{2} (i | s | r) \langle \psi | r \rangle) \cdot \sum_{k=1}^{m} P_{ij}^k \right] \]

Electrostatic potential
\[ \phi(\vec{r}) = \phi_{\text{mol}}(\vec{r}) - \phi_{\text{el}}(\vec{r}) \text{ with } \phi_{\text{mol}}(\vec{r}) = \sum_{\alpha=1}^{N} \frac{Z_{\alpha}}{|\vec{R}_{\alpha} - \vec{r}|} \text{ and } \phi_{\text{el}}(\vec{r}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \int \frac{\sum_{\alpha=1}^{m} P_{ij}^k \cdot \varphi_i(\vec{r}') \cdot \varphi_j(\vec{r}')}{|\vec{r}' - \vec{r}|} d\vec{r}' \]

Dipole moment
\[ \mu = -e \cdot \int \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi_i(\vec{r}) \cdot \varphi_j(\vec{r}) \cdot \sum_{k=1}^{m} P_{ij}^k \right) \cdot \vec{r} \cdot d\vec{r} + e \cdot \sum_{\alpha=1}^{N} Z_{\alpha} \cdot \vec{R}_{\alpha} \]
Table 3. Comparison of the Results Obtained for the Protein Crambin with the ADMA Approach Using Different Surroundings and with Direct Calculations Using Gaussian 98.

<table>
<thead>
<tr>
<th>Method/Size of Surroundings</th>
<th>Hartree–Fock Energy (Hartree)</th>
<th>Error (Hartree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian 98</td>
<td>-17775.2213217</td>
<td></td>
</tr>
<tr>
<td>ADMA/3.0 Å</td>
<td>-17774.5270993</td>
<td>0.6942224</td>
</tr>
<tr>
<td>ADMA/4.0 Å</td>
<td>-17775.0920564</td>
<td>0.1292653</td>
</tr>
<tr>
<td>ADMA/5.0 Å</td>
<td>-17775.1879946</td>
<td>0.0333271</td>
</tr>
<tr>
<td>ADMA/6.0 Å</td>
<td>-17775.2042764</td>
<td>0.0170453</td>
</tr>
<tr>
<td>ADMA/7.0 Å</td>
<td>-17775.2128686</td>
<td>0.0084531</td>
</tr>
<tr>
<td>ADMA/8.0 Å</td>
<td>-17775.2161553</td>
<td>0.0051664</td>
</tr>
<tr>
<td>ADMA/9.0 Å</td>
<td>-17775.2180888</td>
<td>0.0032390</td>
</tr>
<tr>
<td>ADMA/10.0 Å</td>
<td>-17775.2191464</td>
<td>0.0021753</td>
</tr>
<tr>
<td>ADMA/11.0 Å</td>
<td>-17775.2197924</td>
<td>0.0015293</td>
</tr>
<tr>
<td>ADMA/12.0 Å</td>
<td>-17775.2205661</td>
<td>0.0007556</td>
</tr>
</tbody>
</table>
ADMA method test result
Adjustable Density Matrix Assembler (ADMA) for *ab initio* quality macromolecular quantum chemistry

As the example of Crambin shows,

an error of less then 0.5 kcal/mole,
that is, better then *part-per-billion accuracy*

has been obtained with the ADMA method for a

protein,

relative to the conventional Hartree-Fock method.
ADMA method references
Adjustable Density Matrix Assembler (ADMA)
for ab initio quality macromolecular quantum chemistry

The numerical method MEDLA, ancestor of the analytical method ADMA:


ADMA method and some applications:


Some examples for

*ab initio* accuracy

macromolecular quantum chemistry results

using the MEDLA or ADMA linear scaling methods


Proto-Oncogene Tyrosine Kinese 1ABL 0.1 a.u. MIDCO (Molecular Isodensity Contour)
Proto-Oncogene Tyrosine Kinese 1ABL 0.01 a.u. MIDCO
Deoxy-hemoglobin

ADMA
*ab initio*
quality
MIDCO
at 0.1 a.u.
Oxy-hemoglobin

ADMA
*ab initio*
quality
MIDCO
at 0.1 a.u.
Carbon-monoxy-hemoglobin

ADMA
*ab initio*
quality
MIDCO
at 0.1 a.u.
Deoxy-hemoglobin closeup

ADMA

*ab initio*

quality

MIDCO at 0.1 a.u.
Oxy-hemoglobin closeup

ADMA
*ab initio*
quality
MIDCO
at 0.1 a.u.
Carbon-monoxy-hemoglobin closeup

ADMA *ab initio* quality MIDCO at 0.1 a.u.
ADMA Applications in Crystallographic Structure Refinement

One type of application:

P.G. Mezey,
A Crystallographic Structure Refinement Approach Using \textit{Ab Initio} Quality Additive Fuzzy Density Fragments,
ADMA (Adjustable Density Matrix Assembler) references


Natural molecular fragments, functional groups, and holographic constraints on electron densities†

Paul G. Mezey*

Received 23rd January 2012, Accepted 27th April 2012
DOI: 10.1039/c2cp40237h

One of the tools of the shape analysis of molecular electron densities, the Density Threshold Progression Approach used in Shape Group studies can also serve as a criterion for the selection of “natural” molecular fragments, relevant to functional group comparisons, reactivity studies, as well as to the study of levels of relative “autonomy” of various molecular regions. The relevance of these approaches to the fragment-based studies of large molecules, such as biopolymers and nanostructures is emphasized, and the constraints represented by the holographic electron density theorem to this and alternative recent fragment approaches are discussed. The analogies with potential energy hypersurface analysis using the Energy Threshold Progression Approach and connections to level set methods are discussed, and the common features of these seemingly distant problems are described.
3. Another utilization of the holographic principle
Macromolecular Density Matrix Extrapolation (DME) for Macromolecular Conformation Analysis
Brief summary of Density Matrix Extrapolation (DME) and the Löwdin-Inverse-Löwdin (LIL) methods
Holographic Electron Density Theorem,
(a strengthening of the Hohenberg-Kohn theorem):

For a non-degenerate ground state, the complete molecular information is encoded in any small positive volume part of the electron density cloud. Each part “knows” everything about the whole!


One extension for Latent Properties:


For a ground state molecule, a property of one of its excited states is a latent property.

For conformation A, a property of another conformation B is one of the latent properties of A!

A tool for potential surface extrapolation!
OLD IDEA:

Outline of the principle and motivation of extrapolation on potential surfaces, but suggesting a different implementation:

P.G. Mezey: "POTENTIAL ENERGY HYPERSURFACES"
Theoretical background of electron density extrapolation:

Latent Property Holographic Theorem

One relevant latent property: a property of the molecule in a different conformation

Natural Fragment conditions:

How to obtain an extrapolated approximate density matrix

\[ P(x,[x(i)]) \] at conformation \( x \)

from a density matrix \( P(x(i)) \) at conformation \( x(i) \)?

\[ P(x,[x(i)]) = S(x)^{-1/2} S(x(i))^{1/2} P(x(i)) S(x(i))^{1/2} S(x)^{-1/2} \]
Idempotency preservation in the LIL process
The Löwdin transform – inverse Löwdin transform (LIL – transform) for generating an extrapolated approximate density matrix $P(x,[x(i)])$ at conformation $x$ from a density matrix $P(x(i))$ at conformation $x(i)$ is given as

$$P(x,[x(i)]) = S(x)^{-1/2} S(x(i))^{1/2} P(x(i)) S(x(i))^{1/2} S(x)^{-1/2}$$
The **density matrix** \( P \) of a molecule contains information equivalent to the molecular wavefunction.

For the overlap matrix \( S \) and density matrix \( P \).

For proper electronic charge preservation, and for other quality reasons, the so-called **“idempotency condition”** must be fulfilled.

In general, the idempotency condition

\[
PSP = P
\]

can often be used more efficiently in a form reduced to zero:

\[
PSP - P = 0 \quad [X]
\]

where \( 0 \) is the zero matrix.
• How can this be made useful?
The case of non-idempotent approximate density matrices within a conformational domain.

Equation (X), however, is useful even if the original density matrix $P(x(i))$ itself is only an approximate density matrix that is not idempotent.

In such cases of non-idempotent approximate density matrices $P$, however, the $PSP - P$ matrix is no longer the zero matrix 0,

$$PSP - P \neq 0$$
In such a cases of non-idempotent approximate density matrices \( \mathbf{P} \), however, the \( \mathbf{PSP} - \mathbf{P} \) matrix is no longer the zero matrix \( \mathbf{0} \),

\[
\mathbf{PSP} - \mathbf{P} \neq \mathbf{0},
\]
rather, it can be regarded as a matrix representing the deficiency of idempotency,

\[
\mathbf{D} = \mathbf{PSP} - \mathbf{P},
\]
that leads to the scalar measure of **idempotency deficiency**

\[
q(\mathbf{D}) = \text{trace} \left( \mathbf{DD}' \right)
\]
where \( \mathbf{D}' \) is the transpose of matrix \( \mathbf{D} \).
This scalar measure of idempotency deficiency is the sum of the squares of all elements of $D$, and is evidently equal to the scalar zero for perfectly idempotent density matrices $P$:

$$q(D) = \text{trace } (DD^\prime) = 0$$

if

$$D = 0.$$
Using the above general concepts and definitions, the matrix on the left-hand side of eq. (X), ideally a zero matrix, can be regarded as an **Idempotency Deficiency Matrix** $D(P(x(i)))$ when applied to the original density matrix $P(x(i))$, 

$$D(P(x(i))) = \begin{bmatrix} P(x(i)) & S(x(i)) & P(x(i)) - P(x(i)) \end{bmatrix}$$
By the same convention, the **Idempotency Deficiency Matrix** $D(P(x,[x(i)]))$ for the new density matrix $P(x,[x(i)])$ is given as

$$D(P(x,[x(i)])) = P(x,[x(i)]) S(x) P(x,[x(i)]) - P(x,[x(i)])$$
Using these notations, substitution into the LIL eq. shows that the LIL transform of the Idempotency Deficiency Matrix \( D(P(x(i))) \) of original density matrix \( P(x(i)) \), is in fact the Idempotency Deficiency Matrix \( D(P(x,[x(i)])) \) for the new density matrix \( P(x,[x(i)]) \),

\[
D(P(x,[x(i)])) = S(x)^{-1/2} S(x(i))^{1/2} D(P(x(i))) S(x(i))^{1/2} S(x)^{-1/2}
\]
Consequently, the “quality” of the transformed density matrix in terms of “purity”, that is, in terms of the level of satisfying the idempotency condition can be expressed in terms of the “quality” of the original density matrix in terms of its “purity”, that is, in terms of its level of satisfying the idempotency condition, and the relation between these two quality descriptors is provided by the LIL transform.
Conclusion:

the quality of “approximate purity” is maintained within the constraints of the LIL transform
The advantages of density matrix extrapolation and the associated approximate mapping of potential surface domains for quick approximate quantum chemical results are more relevant for larger molecules, rather than for small molecules where direct and more accurate computations are inexpensive.
For illustrative purposes various conformations of the small molecules of ethanol and isomeric dimethyl ether are used here, all belonging to the potential energy surface of the $\text{C}_2\text{H}_6\text{O}$ stoichiometric family.

Three conformers for each of ethanol and dimethyl ether are used as examples. For ethanol, two stable conformers, A and B, as well as the mirror image of B, denoted by $\text{B}'$, are shown, using the 0.1 a.u. isodensity contours, and the analogous isocontours are also shown for three stable conformers of dimethyl ether, denoted by C, D, and E.
The 0.1 a.u. isocontours of structures A, B, B', C, D, and E of the potential surface of the C2H6O stoichiometric family.
For minor rearrangements, such as those among A, B, and B’, on the one hand, and also structures C, D, and E, on the other hand, the density matrix extrapolation provides not only reasonably good estimates but also idempotency preservation. However, extrapolations between these two groups are less accurate. The idempotency deficiency has greater role if the structural differences are large, such as those between ethanol and dimethyl ether.
A special case is illustrated, using mirror image structures of ethanol B and B’, and also showing the 0.1 isodensity contour for an artificial construction, the average electron density of the two enantiomers. In the case of such chiral structures, the deviation between the two electron densities can be quite significant, that also manifests itself in their average having a highly different shape from either of the two enantiomers.
0.1 a.u. isocontours of ethanol B, mirror image B', and their average
In the actual example, however, these deviations are not very severe, and the LIL approach is also viable; note, however, that this is not always the case. Also note that, generating averages for electron densities is even less expensive than the LIL process, and such artificial constructions, in the case of more complicated rearrangements than the actual one shown, can provide a quick test and indication of the expected success of the LIL method.
It should be emphasized, that even if the level of idempotency deficiency is not expected to increase in the LIL process, neither for small nor for large molecules, nevertheless, if a higher level of reliability is needed, than the extrapolated density matrix at any new nuclear arrangement can always be purified. This is especially economical for sparse density matrices.
In the purification of large density matrices obtained from fragment density matrix approaches of macromolecules, such as the ADMA method, the actual matrices are rather sparse, and the efficient eigenvalue methods designed for extremely sparse matrices are advantageous. In such cases, the **Diophantine one-step, non-iterative density matrix purification method** is recommended:

Some of the idempotency-preserving properties of the Löwdin transform – inverse Löwdin transform approach for density matrix extrapolation in the nuclear configuration space are extended to cases where only approximate idempotency is present, as such cases have increased significance with the increased usage of fragment-assembled density matrix methods.
“ADMA – LIL”

A Macromolecular Conformational Analysis Method

Addressing both the large number of electrons and the high dimensionality of potential surface problems
“Star-Path” ADMA-LIL method. In searching for a target conformation, such as a local energy minimum of a macromolecule, an advantageous compromise: use a quick, but less accurate method in an early, crude stage of the search, testing many conformations, follow by a few, more accurate but more expensive computations near the crudely-located target conformation to complete the search. One such approach is a combination of the macromolecular Adjustable Density Matrix Assembler, ADMA, an ab initio quality linear scaling method, and the rather quick Löwdin-Inverse-Löwdin (LIL) density matrix extrapolation method, both involving the same type of density matrices on a common AO basis, ensuring perfect ADMA – LIL compatibility.

Starting with some initial, ab initio quality density matrix, a quick and detailed search of the conformational space can be performed by the LIL density matrix extrapolation method, leading to a new candidate conformation to be recalculated by the more accurate, ab initio quality ADMA method. These steps can be repeated in an iterative fashion, combining the advantages of speed and accuracy.
An analogy for the combination of the fast LIL extrapolation method and the more accurate ADMA method
It would be nice to have first a quick, perhaps not very accurate, but broad “view from above” over a larger part of the energy landscape, and after this quick view, if we see something promising, like a possible valley bottom or a possible mountain pass, we could get down from above, and take a closer look, to get more accurate information.

This would be like taking a quick look from a star up in the sky, and after seeing something promising from above, getting down to Earth to take a closer look for a more precise determination.
Follow a “star-path” for a survey from above,

and when it appears justified,

come down to Earth for a more accurate investigation.

We need two, compatible methods, one fast, and one accurate.
Furthermore, it would be nice to learn not only the energy, but all information about all possible molecular properties of the conformations, and to have this in both cases, from the “star path” far above (perhaps less accurately), and also after coming down to Earth (for more accurate information).

One possibility, not recommended: *empirical potential functions* can provide quick, not very accurate energy information, but no information on details of electron density, or dipole moment, or electrostatic potential, etc. It requires some very different computational methodologies to obtain such additional information, after the empirical potential function approach has found some promising conformation.
It is advantageous, if both the “star path” view, and the “down to Earth” close look can provide information on all molecular properties, within a common framework.

Such a common framework of a “star path” and “down to Earth” combined approach, covering all molecular properties is possible.

The LIL extrapolation method and the ADMA method are fully compatible, with easy information transfer between the two, LIL fast, efficient, but less accurate “from above”, ADMA slower, but of “ab initio” quality.
The Löwdin transform – inverse Löwdin transform (LIL)

How to obtain an extrapolated approximate density matrix

\[ P(x, [x(i)]) \] at conformation \( x \)

from a density matrix \( P(x(i)) \) at conformation \( x(i) \)?

\[ P(x, [x(i)]) = S(x)^{-1/2} S(x(i))^{1/2} P(x(i)) S(x(i))^{1/2} S(x)^{-1/2} \]
The two forms of the holographic electron density theorems are the basis for both the

**Density Matrix Extrapolation, DME methods**

**Additive Fuzzy Density Fragment, AFDF macromolecular methods**

*Some additional motivation, QM/MM *in space, QM/MM *in time:*

In QM/MM, QM and MM are too different, find something in between:
Replace MM with DME, \( \text{DME} = \text{Density Matrix Extrapolation} \)
Replace QM/MM with QM/DME

Practical tool of DME:

**Löwdin-Inverse-Löwdin (LIL) trasformation of density matrices**


**One safeguard: LIL is density matrix “purity-preserving”:**

On the inherited “purity” of certain extrapolated density matrices

Paul G. Mezey *

Scientific Modelling and Simulation Laboratory (SML), Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, 283 Prince Philip Drive, St. John’s, NL, Canada A1B 3X7

ABSTRACT

The method of Löwdin transform – inverse Löwdin transform is extended to approximate density matrix extrapolation in cases where the reference density matrix in not “pure” that is, if it fails the test of idempotency. It is shown that a matrix expression provides a scalar measure of the overall level of impurity and for large, sparse matrices of approximate macromolecular fragment-based quantum chemistry methods the one-step Diophantine density matrix purification method is recommended if a charge-preserving improvement of the extrapolated density matrix is needed.

© 2012 Elsevier B.V. All rights reserved.
A problem of special importance in macromolecules:

By the simplest classical model that still conveys the essential aspects, the intramolecular interactions grow by the square of the formal molecular volume. Some of these interactions are rather strong, such as those along formal chemical bonds, but weaker interactions, conveyed by ranges of lower electron density in between the formal chemical bonds also have an important combined effect.

For quasi-globular macromolecules, such as most proteins, the convex hull of the nuclei surround a region, which can be regarded as the formal “interior” of the macromolecule. A large number of chemical bonds fall within this interior region, but there is also a significant volume of local ranges where the electron density is lower than those found along chemical bonds. These low density ranges, simply by their large relative volume, have important contribution to the interactions within the interior, and have influence on the stability of macromolecular conformations.
The experimental study of low electron density ranges of macromolecules is very challenging and not fully solved, however, macromolecular quantum chemistry may provide useful information. In the quantum chemical settings, the “Low Density Glue” modeling approach [3-6] has been introduced about 15 years ago to the size-dependent importance of the locally weak, but when combined, very extensive component of some of the low-density interactions, which are likely to have a major role in the study of actual shapes and interactions of globular macromolecules.

With focus on the “Low Density Glue”, some of the “holographic” consequences on the computational methodologies will be discussed, including the ADMA [7-12] macromolecular “fuzzy” fragment density matrix methods, and the extrapolation-enhanced LIL-ADMA [13] macromolecular conformation analysis method.

According to one suggestion, generating a series of macromolecular representations using increasing fragment size in ADMA, an extrapolation provides estimates for the expected property values for the ADMA limit, representing the full holographic interdependence of macromolecular parts.
An Approach to Functional Groups in Quantum Chemistry


Some background:

Molecular Isodensity Contours, MIDCOs

A Molecular Isodensity Contour, MIDCO \( G(K,a) \) is defined for a given nuclear configuration \( K \) and electron density threshold \( a \) as the collection of all points \( r \) where the electron density \( \rho(K,r) \) has a value equal to \( a \)

\[
G(K,a) = \{ r : \rho(K,r) = a \}. 
\]
Density Domains and Functional Groups
an Approach Based on Molecular Isodensity Contours, MIDCOs

Density Domain $DD(K,a)$ is a domain enclosed by a MIDCO $G(K,a)$:

$$DD(K,a) = \{ \mathbf{r} : \mathbf{r}(K,r) \geq a \}.$$

The density domain and the fuzzy electron density fragmentation approaches have been suggested for a quantum chemical criterion and representation of formal functional groups.


A useful analogy: consider two molecules near one another.

As long as these molecules have separate identity, each must have some Density Domain containing all the nuclei of the molecule, but none of the nuclei of the other molecule. Separate identity is manifested by such density domains.
Functional groups and limited autonomy within molecules

Two neighbor molecules, A and B: both have MIDCOs enclosing all their nuclei, but not those of the other molecule. This indicates some degree of autonomy for both molecules.
Consider now a *single molecule* and one of its connected density domains $\text{DD}(K,a)$ and the nuclei enclosed by it. The very fact that this subset of the nuclei of the molecule is separated from the rest of the nuclei by the boundary $G(K,a)$ of the density domain $\text{DD}(K,a)$ indicates that these nuclei, together with the local electronic density cloud surrounding them, represent a sub-entity of the molecule, with limited autonomy, and some degree of individual identity.

It is natural to regard such a density domain $\text{DD}(K,a)$ as a criterion and the fuzzy density fragment for the nuclei within $\text{DD}(K,a)$ as a representative of a formal *functional group*. 
Functional groups and limited autonomy within molecules:
use the same condition as for two separate molecules.

*Take a single molecule:*  
The existence of a MIDCO separating a set of nuclei from the rest of the nuclei of the molecule, indicates a *local, limited autonomy* of a Functional Group
Functional groups and limited autonomy within molecules, using the same condition as for two separate molecules.

*Fuzzy functional group*: the fuzzy ADMA fragment for the nuclei of DD(K,a)
Functional groups and limited autonomy within molecules, using the same condition as for two separate molecules.

This criterion does not directly address another aspect of functional groups: reactivity, although it is a natural expectation that functional groups show characteristic reactivity properties.
Allyl alcohol, traditional structural formula
Allyl alcohol, three nested isodensity contours, of 0.2, 0.1, and 0.01 a.u.
Figure 2.5 Some of the high density threshold density domains of the most stable conformation of allyl alcohol, CH$_2$=CH-CH$_2$-OH, as calculated with the GAUSSIAN 90 and GSHAPE 90 programs, using a 6-31G* basis set.
Figure 2.6 Some of the low density threshold density domains of the most stable conformation of allyl alcohol, \( \text{CH}_2=\text{CH}-\text{CH}_2-\text{OH} \), as calculated with the GAUSSIAN 90 and GSHAPE 90 programs, using a 6-31G* basis set.
Functional groups and limited autonomy within molecules: example of ethanol (ethyl alcohol).
No methyl group in ethanol

Example:

According to the quantum chemical definition of functional groups, the ethanol molecule contains the -CH$_2$-CH$_3$ and -OH functional groups, but not the -CH$_3$ functional group.

There is no methyl group as functional group in ethanol!
In ethanol, (ethylalcohol), CH$_3$CH$_2$OH, there is no methyl group, as functional group!
Relations between real molecules through abstract molecules: the reference cluster approach

Paul G. Mezey¹,²,³

Received: 15 June 2015 / Accepted: 13 September 2015 / Published online: 16 October 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Replacing integer atomic numbers representing nuclear charges by continuous variables has already provided some rigorous quantum chemical relations between real molecules, using a formal interpolation through non-physical abstract molecules of continuously varying nuclear charges. Extending this approach to the more general “universal molecule” model, where all discrete parameters of molecules are generalized and replaced by continuous variables, provides further relations, actually interconnecting all real molecules through abstract, non-physical “molecules,” involving, for example, non-integer number of electrons. One simplifying idea of this model is the so-called “reference cluster,” originally defined for isoelectronic systems of a fixed number N of nuclei, where each nucleus is replaced by a possibly fictive nucleus with a nuclear charge equal to the average of the N nuclear charges. Based on the nucleic charge space relation, leading to the electronic energy inequalities, the reference cluster is further generalized here.

Keywords Nuclear charge space · Electronic energy inequalities · Reference cluster · Universal molecule model

1 Introduction

Originally motivated by the United Atom studies of Thirring, Narnhofer, and Lieb [1–3], the idea of exploiting nuclear charge convexity relations between electronic energies of molecules, without involving the extreme case of uniting all nuclei of the molecule into a single, formal nucleus of the United Atom, has lead to a variety of rigorous quantum chemical energy inequalities for various sets of molecules, starting with simple diatomic cases [4, 5].

After the very first examples of such electronic energy inequalities [4, 5], including one of the simplest cases, the Narnhofer-Götz nucleic charge space relation...
Replacing integer atomic numbers **representing nuclear charges by continuous variables** has already provided some rigorous quantum chemical relations between real molecules, using a formal interpolation through nonphysical abstract molecules of continuously varying nuclear charges.

Extending this approach to the more general “universal molecule” model, where all discrete parameters of molecules are generalized and replaced by continuous variables, provides further relations, actually interconnecting all real molecules through abstract, non-physical “molecules,” involving, for example, non-integer number of electrons. One simplifying idea of this model is the so-called “reference cluster,” originally defined for isoelectronic systems of a fixed number \( N \) of nuclei, where each nucleus is replaced by a possibly fictive nucleus with a nuclear charge equal to the average of the \( N \) nuclear charges. Based on the earlier results, some new relations are derived interrelating energies and some other properties of some real molecules, and also providing a unifying framework for the utilization of both symmetry and energy relations of the universal molecule model.
Compensation Effects in Molecular Interactions and the Quantum Chemical Le Chatelier Principle

Paul G. Mezey*

Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, 283 Prince Philip Drive, St. John’s, NL A1B 3X7, Canada

ABSTRACT: Components of molecular interactions and various changes in the components of total energy changes during molecular processes typically exhibit some degrees of compensation. This may be as prominent as the over 90% compensation of the electronic energy and nuclear repulsion energy components of the total energy in some conformational changes. Some of these compensations are enhanced by solvent effects. For various arrangements of ions in a solvent, however, not only compensation but also a formal, mutual enhancement between the electronic energy and nuclear repulsion energy components of the total energy may also occur, when the tools of nuclear charge variation are applied to establish quantum chemically rigorous energy inequalities.

INTRODUCTION

Among molecular interactions, solvent—solute interactions are ranking very high in importance. Chemical reactions occurring in solutions are fundamentally affected by such interactions, consequently, they have primary roles in a range of chemical fields, including biochemical processes, and also chemical industry. Understanding and modeling such interactions and the reactions occurring in solutions represent many challenges. For the energy inequalities derived in this study, the presence of a solvent is essential, and in this context, the main effect of the solvent is making the geometrical arrangements considered for these ions realistic. The nuclear arrangements of the solute ions are restricted, nevertheless, they still have some freedom: two distance parameters, $D$ and $d$ of the model have only some lower limits, otherwise they can be chosen rather freely, allowing a whole range of geometries. This is similar to one feature of the polarizable continuum model of solvents, where...
Following a somewhat simplified version of the original proof in ref 24, consider three isoelectronic molecular systems, \( M^{(1)} \), \( M^{(2)} \), and \( M^{(3)} \) in their electronic ground states, where the nuclear locations match, but the nuclear charges can be different, and even zero nuclear charges, that is, "dummy nuclei", are allowed. Note that each of these systems may be an actual set of ions, or an arrangement of several molecules.

Let us denote the corresponding nuclear charge vectors by \( z^{(1)} \), \( z^{(2)} \), and \( z^{(3)} \), respectively, where the ordering of the nuclei as components of these vectors reflects the common nuclear geometry \( r \). If \( z^{(3)} \) is a convex combination of \( z^{(1)} \) and \( z^{(2)} \),

\[
z^{(3)} = \alpha z^{(1)} + (1 - \alpha) z^{(2)}
\]

(6)

where

\[
0 \leq \alpha \leq 1
\]

(7)

then the same must apply for their corresponding electronic Hamiltonians linear in \( z \):

\[
H_e(z^{(3)}, r) = \alpha H_e(z^{(1)}, r) + (1 - \alpha) H_e(z^{(2)}, r)
\]

(8)

If the electronic wave function of system \( M^{(j)} \) is denoted by \( \Psi_e^{(j)} \), then on the basis of eq 8, the electronic energy expectation value of system \( M^{(3)} \) can be written as
\begin{align*}
  \langle \Psi^{(3)}_e | H_e(z^{(3)}, r) | \Psi^{(3)}_e \rangle & = \alpha \langle \Psi^{(3)}_e | H_e(z^{(1)}, r) | \Psi^{(3)}_e \rangle + (1 - \alpha) \langle \Psi^{(3)}_e | H_e(z^{(2)}, r) | \Psi^{(3)}_e \rangle \\
  | \Psi^{(3)}_e \rangle
\end{align*}

According to the variational theorem, the right-hand side cannot increase if, for each of the two Hamiltonians, \( H_e(z^{(1)}, r) \) and \( H_e(z^{(2)}, r) \), the corresponding wave functions, \( \Psi^{(1)}_e \) and \( \Psi^{(2)}_e \) are used, respectively, instead of the common wave function \( \Psi^{(3)}_e \). Therefore,

\begin{align*}
  \langle \Psi^{(3)}_e | H_e(z^{(3)}, r) | \Psi^{(3)}_e \rangle & \geq \alpha \langle \Psi^{(1)}_e | H_e(z^{(1)}, r) | \Psi^{(1)}_e \rangle + (1 - \alpha) \langle \Psi^{(2)}_e | H_e(z^{(2)}, r) | \Psi^{(2)}_e \rangle \\
  | \Psi^{(2)}_e \rangle
\end{align*}

that is,

\begin{equation}
  E^{(3)}_e \geq \alpha E^{(1)}_e + (1 - \alpha) E^{(2)}_e
\end{equation}

The above inequality relation is rigorously valid, typically, with a substantial margin.
We shall apply the electronic energy inequality, eq 11, for three sets of rather general systems of ions, shown in Figures 1, 2, and 3, where we shall consider several actual realizations of these general models. In all diagrams, the charges of ions X, Y, and A have the same sign, whereas the charge of ion Q has the opposite sign. In the actual applications, we shall use both anions and cations, both singly and doubly charged, but following the patterns specified in these figures.

In each of the three figures, three families of four ions are shown, where each family is arranged as two ion pairs, and the energy relations are derived between these families. The energy relations hold when the distance between the members of each pair is increased from d to D and from D back to d. The figures show how the energy changes as the distance between the ions changes.
Three ionic systems, $M_1$, $M_2$, and $M_3$ of identical local geometries in solution, arrangement (b):

$$M^{(1)}: \quad Q \quad X \quad Y \quad Q$$
$$M^{(2)}: \quad Q \quad Y \quad X \quad Q$$
$$M^{(3)}: \quad Q \quad A \quad A \quad Q$$

Figure 2. Arrangement (b) of three isoelectronic sets of ions.

Three ionic systems, $M_1$, $M_2$, and $M_3$ of identical local geometries in solution, arrangement (c):

$$M^{(1)}: \quad X \quad Q \quad Q \quad Y$$
$$M^{(2)}: \quad Y \quad Q \quad Q \quad X$$
$$M^{(3)}: \quad A \quad Q \quad A \quad Q$$

Figure 3. Arrangement (c) of three isoelectronic sets of ions.
\[ E_e^{(3)} \geq 0.5E_e^{(1)} + 0.5E_e^{(2)} \] (21)

Specifically, if the three arrangements of Figures 1–3 are applied to the first set, the general electronic energy inequality (21) becomes the following three actual inequalities, where in the figures, \( \text{Cl}^-\), \( \text{I}^-\), \( \text{Br}^-\), and \( \text{Na}^+\) are taking the roles of the general ion symbols \( X\), \( Y\), \( A\), and \( Q\), respectively:

**Figure 1:**

\[ E_e(\text{Br}^-\text{Na}^+ \ldots \text{Br}^-\text{Na}^+) \]
\[ \geq 0.5E_e(\text{Cl}^-\text{Na}^+ \ldots \text{I}^-\text{Na}^+) + 0.5E_e(\text{I}^-\text{Na}^+ \ldots \text{Cl}^-\text{Na}^+) \] (22)

**Figure 2:**

\[ E_e(\text{Na}^+\text{Br}^- \ldots \text{Br}^-\text{Na}^+) \]
\[ \geq 0.5E_e(\text{Na}^+\text{Cl}^- \ldots \text{I}^-\text{Na}^+) + 0.5E_e(\text{Na}^+\text{I}^- \ldots \text{Cl}^-\text{Na}^+) \] (23)

**Figure 3:**

\[ E_e(\text{Br}^-\text{Na}^+ \ldots \text{Na}^+\text{Br}^-) \]
\[ \geq 0.5E_e(\text{Cl}^-\text{Na}^+ \ldots \text{Na}^+\text{I}^-) + 0.5E_e(\text{I}^-\text{Na}^+ \ldots \text{Na}^+\text{Cl}^-) \] (24)