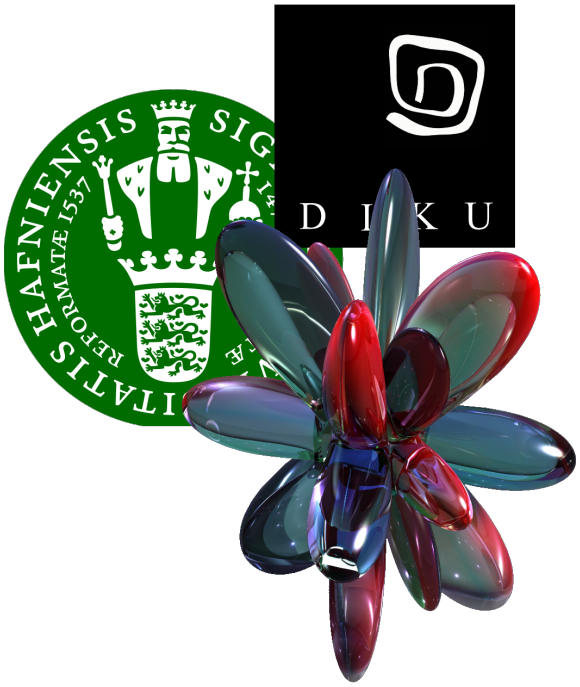
JAMES AVERY
(avery@diku.dk)

NEW COMPUTATIONAL METHODS IN THE QUANTUM THEORY OF NANO-STRUCTURES

SUPERVISED BY STIG SKELBOE



PH.D. DISSERTATION

DIKU
DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF COPENHAGEN
THE PHD SCHOOL OF SCIENCE, FACULTY OF SCIENCE
DENMARK, JANUARY 2011

Dedicated to my loving parents, Birthe and John Avery.

Abstract

This thesis presents three new methods for the computation of electronic structure problems in the quantum theory of nano-structures:

1. An extension of the linearly scaling *SIESTA* DFT-method has been developed to employ the finite element method, and is implemented in software. The aim is to be able to solve electronic structure problems for quantum mechanical systems with hundreds to many thousand electrons *ab initio*, embedded in a large environment that is treated using classical electrostatics. As an example, we calculate properties of a molecular single electron transistor, in which a large organic molecule interacts with metallic and dielectric regions.
2. Two schemes have been developed (and implemented) for the rapid evaluation of multicenter interelectron repulsion integrals for exponential type orbitals (ETO's). This is a long standing problem in computational quantum theory, but one that is rarely attacked. The present work provides solutions for the special type of ETO's called *Coulomb Sturmians*.
3. A method has been developed for solving molecular systems using isoenergetic many-electron molecular basis functions constructed from many-center Coulomb Sturmians. The automatic scaling properties of Sturmian basis functions yields high flexibility and accuracy, and allows large parts of calculations to be done off-line. The method is implemented in part, and calculations are performed for diatomic molecules.

Sammenfatning

I denne afhandling præsenteres tre nye metoder til kvantemekaniske beregninger af nanosystemers elektronstruktur:

1. Vi har udvidet den lineært skalerende DFT-metode *SIESTA* til at anvende finite elementmetoden, og har udviklet software som implementerer den udvidede fremgangsmåde. Målet er at kunne løse elektronstrukturproblemer for kvantemekaniske systemer i størrelsesorden fra hundreder til mange tusinde elektroner. Det kvantemekaniske system kan være indlejret i et stort ydre miljø, som behandles med klassisk elektrostatik. Den udviklede software benyttes til at beregne egenskaber for en molekylær enkeltelektrontransitor (*single electron transistor*), i hvilken et stort organisk molekyle interagerer med metalliske og dielektriske regioner.
2. Vi har udviklet og implementeret to forskellige fremgangsmåder til effektivt at udregne molekylære elektronfrastødningsintegraler for eksponentialorbitaler. Dette er et problem for kvantemekaniske beregninger, som går flere årtier tilbage, men som sjældent angribes. I denne afhandling præsenteres løsninger for en bestem type eksponentialorbitaler: *Coulomb Sturm-baser*.
3. Vi har udviklet og implementeret en metode til beregning af molekylære systemer ved brug af isoenergiske molekylære mangeelektron-basisfunktioner, som er bygget af mange-centrum Coulomb Sturm-baser. Automatisk skalering af Sturm-baserne giver automatisk stor fleksibilitet og beregningsnøjagtighed, og tillader os desuden at foretage størstedelen af beregningerne off-line. Metoden er delvist implementeret, og vi har foretaget pilotberegninger på diatomare molekyler.

Preface

Summary of the Thesis

This thesis is submitted in partial fulfillment of the requirements for the Ph.D. degree at the Department of Computer Science, University of Copenhagen (DIKU). The thesis describes work conducted during the period from April 2006 to January 2011, under supervision of Prof., Dr. Techn. Stig Skelboe. It consists of three distinct parts, each describing a major line of my research during the past three years. The first part describes an extension of a linearly scaling[1] DFT method to use the finite element method to give a multi-scale approach, as well as calculations using the new method to calculate properties of a molecular single-electron transistor. The method scales to describe physical systems that embed quantum-active parts in arbitrarily large classically treated devices. Linear scaling of the DFT method ensures that we can treat quantum-active parts with hundreds to thousands of electrons. The second part of the thesis treats an old problem in quantum chemistry: the efficient computation of many-center interelectron repulsion integrals for exponential type orbitals. Two approaches that both take advantage of automatic scaling properties of a particular type of ETO, Coulomb Sturmians, are presented. Finally, the third part of the thesis applies the generalized Sturmian method (Avery and Avery [2006]) to calculations on molecules using isoenergetic configurations of Coulomb Sturmian orbitals. This work is at a very experimental stage, but preliminary results look promising. In addition, five publications that were written during the course of my Ph.D. are included as appendices: Three published papers, one book chapter, and one nearly finished paper that will be polished and published shortly.

Acknowledgements

Three people have shaped the work presented in this thesis:

I would first of all like to thank my supervisor, Professor Stig Skelboe, for encouragement, countless discussions and for his invaluable help in crushing all of the obstacles that have appeared over the course of the last three years.

Secondly, I owe many thanks to Kurt Stokbro from QuantumWise, in collaboration with whom the first part of this thesis was developed. Kurt has functioned as a second supervisor, and had it not been for the many hours of explaining the physical systems and analyzing the many problems that appeared along the way, I would never have been able to complete this work.

Thirdly, I must thank my father, John Avery. Part II and III of this thesis lies in continuation of our ongoing work on the generalized Sturmian method and isoenergetic, automatically scaling basis functions in computational quantum theory. I cannot overstate the joy I derive from our work together. Here is to many more years of knocking loudly on the doors of quantum theory and refusing to leave until it hands over its secrets.

[1] Under certain conditions.

A number of people have contributed to the thesis in less direct, but nonetheless important ways.

Thanks go to Prof. Peter Schwerdtfeger and Matthias Lein at Centre for Theoretical Chemistry and Physics in Auckland, New Zealand, for an enjoyable four-month research visit. I look very much forward to working together in the near future.

To my mother Birthe, and my sisters Anne and Julie: Your love is a great source of strength, which I cannot imagine to be without.

Penultimately, a thank you to the good friends who have been an endless source of encouragement and relentless support, and who have helped keep me sane during the past three years. To my brothers in arms since childhood, Christian Munch, Casper Bøyesen, and Mathias Bertelsen: May our clans rise in the next 50 years as they have the past 25. The world shall be our box of matches, and we shall light each continent ablaze. And my dear friends: Cousin Schmidt, Niller, Lea, Brinja and Anders, Dr. Kaad, the Kyndes and the Trier Troldborgs. Your friendships are of indescribable importance to me.

And finally, very special thanks go to my wonderful girlfriend, the esteemed Dr. Anastasia Borschevsky: brilliant scientist and femme fatale extraordinaire. You are beautiful and clever and not at all fat! I cannot wait to rejoin you in the land of the long white cloud, from where we will rule the South Seas and find many small aquatic creatures to murder and, with great pleasure, consume.

Contents

Introduction

> *"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."*
>
> Dirac [1929]

Since the discovery of the wave equations for matter in 1926, we have possessed a key that in principle unlocks, to paraphrase P.A.M. Dirac, all of chemistry and most of physics. Almost any property that is covered by our current understanding of physics, chemistry, or indeed biology, can theoretically be calculated from first principles by way of a few, very simple, fundamental laws of nature. However, being able to write down the governing equations of matter is an entirely different matter than is solving these equations; The wave equations for interacting particles have known solutions only for a few extremely simple systems. Approximations must therefore be used, and the practical utility of quantum theory depends on discovering approximations that will allow reasonably accurate calculations to be performed on systems that are large enough to be of interest.

Great strides have been made over the past century towards this goal, and applications have showed quantum theory, in the cases where it can be practically employed, to predict experiments to an incredible accuracy, and to yield insights into the workings of natural phenomena that were previously unattainable. But while the advances have been many, solving the Schrödinger or Dirac wave equations, even by approximation, remains a formidable task: For electronic systems, they are partial differential equations of $3N$ spatial variables and N spin variables, and inter-electron interaction prevents separation of variables. Modeling even a single atom is therefore a many-dimensional problem, and the computational complexity grows explosively with the number of particles. The development of evermore efficient approximation methods is thus not so much motivated by making calculations more convenient as it is by making them possible at all. It is an ongoing effort that will see steady progress for many decades to come, each step opening doors to new areas of science that we can treat *ab initio*.

This thesis takes another few steps in the direction of increased efficiency and hence wider applicability of computational quantum mechanics. In it, I describe the three dominant lines of research in my work over the past three years. Correspondingly, the thesis is divided into three distinct parts:

In Part I, I present a computational method that combines a linearly scaling density functional theory model with the finite element method. The aim is to be able to compute properties of nanodevices using a model that combines a quantum mechanical description with a classical electrostatic one. This work was done in collaboration with Kurt Stokbro from Quantumwise A/S, and Kristen Kaasbjerg from the Technical University of Denmark (DTU).

In Part II, I present two methods for efficiently calculating multi-center interelectron repulsion integrals for a special class of exponential type orbitals called Coulomb Sturmians. Despite being the best match for accurate representation of wave functions, exponential type orbitals are rarely used due to the great difficulty in evaluating these integrals. The current work exploits the special automatic scaling properties of Coulomb Sturmians to relegate all heavy computational work to off-line calculations that are performed once and for all.

In Part III, I present a method for solving molecular systems using isoenergetic many-electron molecular basis functions constructed from many-center Coulomb Sturmians. This is a second case where automatic scaling of basis functions aids both in terms of flexibility and accuracy, and in allowing large parts of calculations to be done off-line. This part is a work in progress, but pilot calculations have shown promising results. Both the work in Part II and in Part III was done in collaboration with my father, John Avery.

Part I: A linearly scaling finite-element DFT

Quantum theory is certainly a branch of theoretical science that has more than paid for itself, since insights gained from quantum theory led to the invention of the transistor. These insights came from an understanding of the electron band structure of perfect crystals, especially the narrow band gaps in semiconducting crystals. Here, the systems studied were macroscopic, but quantum theory was aided by the periodic structure of the crystals, and the use of translational symmetry allowed calculations to be done even on crystals that would otherwise be far beyond the reach of quantum theoretical calculations.

The invention of transistors led to an enormous increase in the speed of computers. The limiting factor in computational speed is the distance that electrical signals must travel within the processing units. This is remarkable, since these signals travel at the speed of light, but nevertheless, it is the limiting factor. Hence, the quest for greater speed has become a quest for miniaturization. However, the photo-etch techniques, that are used in the production of transistors are limited in resolution by the wave length of the light used. Thus, for devices smaller than that wave length to be possible, it is necessary to introduce new methods of construction.

Nano-science deals with structures whose dimensions are on the order of the wave-length of ultraviolet light. Hence, such structures involve from around a hundred to many thousands of atoms. Methods for constructing structures on the nano scale draw on insights from biology, where autoassembly of sub-cellular structures is one of the most important principles. Autoassembly requires both spatial and electrostatic complementarity. The hope is that autoassembly can be used to produce transistors in which single molecules will play the role of semiconductors, the charging energies (the ionization energies and electron affinities) playing a role similar to that played by the small band gap in semiconducting crystals. This type of nano-scale transistor is called a *molecular single electron transistor*, or molecular SET.

In order to calculate the charging energies, quantum theory is needed. It is needed also to calculate the excess charge distributions that play a role in autoassembly. However, the size of nano-structures implies that the usual methods of quantum chemistry cannot be used: the structures are too large. Nor is there the translational periodicity which led to the insights on which large-scale transistors are based. To complicate matters still further, the single molecule which is a central part of a molecular SET, is greatly influenced by its environment, which includes both a dielectric, and the gold electrodes to which it is attached. Thus, special methods must be introduced to allow quantum calculations on this large and complicated system.

In this part of the thesis, a linearly scaling DFT method is combined with the finite element method and explored as a possible method for calculating large quantum mechanical systems embedded in a much larger classically treated environment. The methods developed here can be

used whenever a detailed description of the quantum-active part of a nano-structure is needed, combined with a cruder description of its environment. In the method developed in this thesis, the large environment is treated using classical electrostatics, and this is coupled to a quantum theoretical solution for the quantum-active part of the system.

The interaction of the quantum system with the electrostatic environment is based on a model devised by Kristen Kaasbjerg (Kaasbjerg and Flensberg [2008]), developed for the semi-empirical Hückel method. In this model, the total energy function is extended to include interactions with a number of dielectric and metallic regions surrounding the molecular system.

While the finite element method allows us to treat very large classical regions with negligible computational overhead, the choice of DFT-method sets the limit on the size of the quantum-active part of the nano-structures. The aim is to be able to handle electronic systems ranging from hundreds to many thousand electrons: a regime for which ab initio quantum theoretical calculations have become possible, albeit not easy, during the past decade with the advent of linearly scaling methods. The particular method that we have extended is the SIESTA method (Soler et al. [2002]). This method scales linearly in the number of basis functions, provided that the number of overlapping orbitals per spatial point is bounded by a constant as the system is grown. The software employed in this thesis extends an independent implementation of the SIESTA method by Kurt Stokbro and Dan Erik Petersen.

Part II: Interelectron repulsion integrals

The second part of this thesis is devoted to the problem of evaluating multi-center interelectron repulsion integrals (ERIs) with exponential type orbitals (ETOs). The methods developed offer potentially far more accurate methods in quantum chemistry than the ones that are currently in most popular use. The most widely used computer programs in quantum chemistry make use of Gaussian basis sets exclusively despite the fact that these basis sets are intrinsically inferior in accuracy to those based on exponential type orbitals: Electronic wavefunctions are known to decay exponentially at long ranges, a feature that cannot be reproduced faithfully by Gaussians, nor can Gaussians reproduce the exponential cusps that occur at the nuclear sites. Why, then, have Gaussian basis sets been used for the last 60 years? Why do they now dominate the field to point of being nearly exclusive? The reason for this is the mathematical difficulty of evaluating multi-center interelectron repulsion integrals with exponential type orbitals, and the great ease for which this can be done with Gaussians. In a sense, it is the case of searching for our missing keys under the street lamp instead of in the bushes where we lost them: because here, there is light. Because of this, a rapid method for evaluating many-center ERIs with exponential type orbitals could be a small breakthrough in the field.

The approach described in this thesis makes use of a special type of ETO, Coulomb Sturmian basis functions. These functions have the interesting property that their Fourier transforms can be simply related to four dimensional hyperspherical harmonics. The method for evaluating ERIs that is developed in this thesis makes use of a Fourier transformed representation of the charge densities, and it also makes use of the properties of hyperspherical harmonics. The most important cases are treated exactly by this method. For the difficult case of three- and four-center ERIs, we introduce two different approximation schemes, one based on hyperspherical harmonics and Legendre polynomial expansions, and the other on a special form of Gaussian expansion especially adapted to Coulomb Sturmian basis functions. The asymptotic problems with Gaussian expansions are largely avoided, since these integrals vanish rapidly. The particular Gaussian expansion method has the feature that the expansion coefficients are universal, and scale automatically to accommodate both diffuse and contracted orbitals. Because of the automatic scaling, the heavy work of computation can be done off line, and precomputed closed form expressions can be used in general situations. The work has been carried through in detail for the simple case

of diatomic molecules, but a framework for the general case is both presented and implemented.

Part III: Molecular calculations based on Coulomb Sturmians

Finally, the third part of this thesis is devoted to a new method for quantum chemical calculations in which isoenergetic many-electron molecular basis functions constructed from many-center Coulomb Sturmians are used to build up the wave function. The method is closely analogous to the generalized Sturmian method for atoms, described for example in Avery and Avery [2006] and Avery [2008a]. Because of time limitations, only a few very simple examples have been treated, but the results are very promising. Among the features of the method is an automatic adjustment in scaling of the basis functions, so that the same basis set can be used regardless of whether the wave function described is diffuse or contracted. As the interatomic distances are increased or decreased, the basis set adapts automatically to give an appropriate description. Furthermore, the diffuse orbitals of the excited states of the system are appropriately described at the same time as the contracted orbitals of the ground state, because each state is automatically given its own scaling factor.

Part I

A linear-scaling finite-element DFT

CHAPTER 1

Introduction

Nanostructures are small enough so that chemistry and quantum effects are important, yet so large that most quantum theoretical methods are computationally infeasible. For systems involving hundreds to thousands of electrons, we are essentially left with semi-empirical methods and DFT as the only realistic lines of attack.

Conventional DFT-methods scale as $\mathcal{O}(N^3)$ in the number of electrons. Although this is much more efficient than "real" many-body quantum chemical methods, the growth in resource consumption still prohibits practical calculations of medium to large nano-systems. However, there has been a good deal of progress in the last decade in the development of DFT methods that exploit sparseness to produce methods that effectively require $\mathcal{O}(N)$ resources for N-electron problems. One such order-N method is the SIESTA method (*Spanish Initiative for Electronic Simulations with Thousands of Atoms*), described by Soler and coworkers in Soler et al. [2002].

The SIESTA method is based on basis sets of linear combinations of atomic orbitals (LCAO), which is common in quantum theoretical methods. However, they use numerical basis functions with compact support optimized for pseudopotentials that yield very short-range orbitals that are non-zero only in a small region of space. This results in very sparse equations that can be solved with $\mathcal{O}(N)$ time and space for many classes of systems.

The method relies on representing the electron density on a uniform real space grid. Because memory requirements for a uniform grid scale as $\mathcal{O}(L^3)$ in the grid's diameter L, this scheme provides good efficiency only when the region of space represented by the grid is small compared to the number of electrons, or rather when the electron density is spread out more or less evenly over the represented region. This is generally the case when solving for molecules or crystals in vacuum.

However, there are many applications in nano-science for which we would like a detailed quantum theoretical treatment of only a small sub-part of the system that is situated in a much larger environment with which it interacts. Consider for example nano-devices, in which a molecule, or a similar quantum-active nano-structure, resides in a (relatively) macroscopic device that can be treated classically to a good approximation.

The method presented in the following chapters extends the SIESTA method, lifting the real space representation from a regular grid to finite element discretizations of space. This allows us multiple levels of detail in treating the system as a whole: The regions where little interesting quantum chemistry goes on can be treated with classical electrostatics using crude representations, while at the same time the quantum chemical parts are treated with high accuracy. The interaction of the quantum system with the electrostatic environment is based on a model devised by Kristen Kaas-

bjerg (Kaasbjerg and Flensberg [2008]), developed for the semi-empirical Hückel method. In this model, the total energy function is extended to include interactions with a number of dielectric and metallic regions surrounding the molecular system.

Finite Element Methods are ubiquitous in almost all fields of physics and engineering, but are very rarely used in quantum theory. A few authors have done so, however, for example Bylaska et al. [2009], Toda et al. [2009], and Pask and Sterne [2005].

The software that has been developed is comprised of two parts: A general library `libspace` with general support for the described methods, and an extended version of the SIESTA-like DFT-code `qscf`, developed by Dan Erik Petersen and Kurt Stokbro.

CHAPTER 2

The Finite Element Method

The finite element method is among the most commonly employed approximation methods used in physics and engineering, if it is not *the* most commonly employed method. It is used to compute the physical properties of anything from buildings, cars, and airplanes to fluids, gasses or even human organs. However, the finite element method is almost never used in quantum chemistry and -physics.

The method has many strengths, the most important of which are the ease with which problems can be described simultaneously at multiple levels of detail, the ease with which the levels of approximations can be systematically controlled, and finally: a solid theoretical foundation and rigid estimates of both global and local error that can be used to dynamically adapt the approximation to be locally refined in regions where the real solution requires detailed representations and coarse in regions where crude representations suffice. Even though the real solution is in fact unknown, the approximation theory of finite element analysis can tell us how far off we are, in several different norms.

In this chapter, we will only give a brief introduction, but there exists an enormous amount of excellent literature on finite element analysis, covering the whole range from the very theoretical underpinnings to practical applications. The author has especially found Brenner and Scott [2008] and Johnson [1987] extremely useful. Although the material presented in this chapter is not new, the reader already familiar with the finite element method is encouraged to read or at least skim the chapter nonetheless, since important nomenclature is introduced.

2.1 Sobolev spaces

The theory of Sobolev spaces offers a simple but incredibly strong framework for constructing and analyzing approximations to function spaces on manifolds. Sobolev spaces essentially lift the theory of differentiable functions to an L^p-setting[1] through a generalization of the ordinary calculus derivative called the *weak* derivative. Especially useful are the H^k Sobolev spaces, which are subspaces to the Hilbert space L^2 of square integrable functions. This is the natural setting for finite element analysis, and it thus pays to sacrifice a little time to introduce the theory.

[1] We remind the reader that $L^p(\Omega;\mathbb{K})$ is the vector space consisting of functions $f : \Omega \to \mathbb{K}$, for which the integral $\int_\Omega d\mathbf{x}\, |f(\mathbf{x})|^p$ is well-defined and finite. $L^p(\Omega)$ is shorthand for $L^p(\Omega;\mathbb{C})$, $L^p(\Omega;\mathbb{R})$ or either, depending on the context. Most often, the choice makes no difference.

We will use the following shorthand for writing derivatives:

$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$$

where $|\alpha| := \alpha_1 + \cdots + \alpha_n$. For example, $\frac{\partial^3}{\partial x_1^2 \partial x_2} = D^{(2,1)}$.

The *weak derivative* allows us to define the derivative for a much larger class of functions than the ones differentiable by the ordinary derivative known from calculus. Where the calculus derivative requires the limit

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

to be defined in every point x, the weak derivative can be defined instead on the space of *locally integrable functions*, L^1_{loc}. Rather than being point-wise defined, the derivative is defined in terms of the inner products with every member of $C_0^\infty(\Omega)$, i.e. the smooth functions with compact support on the domain Ω. The weak derivative is defined as follows:

Definition 2.1.1 (Weak derivative)**.** *Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $f \in L^1_{loc}(\Omega)$. The* weak derivative *$D^\alpha_w f$ is defined if there exists a function $g \in L^1_{loc}(\Omega)$ such that*

$$\int_\Omega g(x)\phi(x)dx = (-1)^{|\alpha|} \int_\Omega f(x) D^\alpha \phi(x) dx \quad \forall \phi \in C_0^\infty(\Omega)$$

If such a g exists, we define $D^\alpha_w f = g$.

It is a true generalization of the ordinary derivative: Whenever the point-wise derivative is defined everywhere on Ω, the weak derivative coincides with it, i.e. $D^\alpha_w f = D^\alpha f$.

The weak derivative gives rise to the following partially defined function on L^p-spaces:

Definition 2.1.2 (Sobolev norm)**.** *Let $f \in L^1_{loc}(\Omega)$, and $k \geq 0$. If $D^\alpha_w f$ exist, we can define for each $1 \leq p < \infty$*

$$\|f\|_{W^k_p(\Omega)} := \left(\sum_{|\alpha| \leq k} \|D^\alpha_w f\|^p_{L^p(\Omega)} \right)^{1/p}$$

Note that, despite its name, the *Sobolev norm* is not a norm on L^p, but on certain subspaces called *Sobolev spaces*. These are in fact defined as the subspaces on L^p, for which $\|\cdot\|_{W^k_p}$ is a norm:

Definition 2.1.3 (Sobolev space)**.** *A* Sobolev space *is a space of the form*

$$W^k_p(\Omega) := \left\{ f \in L^1_{loc}(\Omega) \;\middle|\; \|f\|_{W^k_p(\Omega)} < \infty \right\}$$

In words, this states that W^k_p is simply the subspace of the usual L^p-space that consists only of functions $f \in L^p$ for which also the weak derivative $D^\alpha f \in L^p$ whenever the degree $|\alpha| \leq k$. In particular,

$$L^p = W^0_p \supset W^1_p \supset W^2_p \supset \cdots$$

We will use only a particular class of Sobolev spaces that are subspaces of L^2 and are thus also Hilbert spaces. These are the spaces

$$H^k(\Omega) := W^k_2(\Omega)$$

with the inner product

$$\langle f|g\rangle_{H^k} := \sum_{|\alpha|\leq k} \langle D_w^\alpha f|D_w^\alpha g\rangle_{L^2} \tag{2.1.1}$$

That is: H^k is the subspace of L^2 consisting of square-integrable functions for which all weak derivatives up to degree k are also square-integrable. These spaces are appropriate for Hilbert space solutions to $2k$-degree differential equations.

We note that the space $C_0^{k-1}(\Omega)$ of at least $(k-1)$-differentiable functions with compact support on Ω is obviously contained in $H^k(\Omega)$. In particular, also $C_0^\infty \subseteq H^k$, and in fact both these space are dense in H^k under the H^k-norm. There are two ramifications arising from this: First, the inclusion implies that if a $2k$-degree differential equation has a classical solution in $C_0^{2k}(\Omega)$, this solution is the same as the weak solution in H^k, since the strong and the weak derivative must then coincide. Second, we can approximate solutions in H^k to any desired degree using only $(k-1)$-differentiable or even smooth functions with compact support.

In the following, we will only consider Sobolev spaces that are function spaces over some subset $\Omega \subseteq \mathbb{R}^n$. In addition, since we will only be solving real valued differential equations, we consider only real valued Sobolev spaces, i.e. with the scalar field $\mathbb{R}$ instead of $\mathbb{C}$. While neither of these restrictions are strictly necessary, they do simplify the exposition, and are general enough for our purposes. However, with a few changes, the results hold for more general domains and complex Sobolev spaces.

2.2 Variational formulation of partial differential equations with boundary conditions

Consider a partial differential operator of order k on $\Omega \subseteq \mathbb{R}^n$:

$$\mathsf{A} = \sum_{|\alpha|\leq k} a_\alpha D^\alpha$$

A acts as a linear operator on the smooth functions $C^\infty(\Omega)$. Consider now the k^{th} order partial differential equation: Find $u \in C^\infty(\Omega)$ such that

$$\mathsf{A}u(\mathbf{x}) = f(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega \tag{2.2.1}$$

For now, we ignore the issue of boundary conditions. By replacing the classical derivative in A by the weak derivative, we can pose the problem in terms of distributions. Let $V \subseteq L^2(\Omega)$ be an appropriate Sobolev space. Given f, find $u \in V$ such that

$$\mathsf{A}u = f$$

By the Riesz Representation Theorem, we can write this in terms of the corresponding linear functionals:

$$\mathsf{A}\,|u\rangle = |f\rangle \tag{2.2.2}$$

or equivalently

$$\forall v: \quad \langle v|\mathsf{A}|u\rangle = \langle v|f\rangle \tag{2.2.3}$$

where the L^2 inner product is used. This allows us to use Hilbert space methods to solve the partial differential equation. If a classical solution exists, we know u will coincide with it, but many equations have only weak solutions that nonetheless have physical significance.

Above, we skirted two important questions: What is "an appropriate Sobolev space" V, and when does Equation (2.2.2) have solutions? This will be made precise in the following.

Definition 2.2.1 (Bilinear form)**.** *A* bilinear form *on the (real) vector space V is a mapping $\mathsf{A} : V \times V \to \mathbb{R}$ that is linear in both of its arguments.*

We say that A *is* bounded *if there exists a constant $C_1 > 0$ such that*

$$|\mathsf{A}(f,g)| \leq C_1 \|f\| \|g\| \quad \forall f,g \in V \tag{2.2.4}$$

It is called elliptic *or* coercive *if there exists $C_2 > 0$ such that*

$$|\mathsf{A}(f,f)| \geq C_2 \|f\|^2 \quad \forall g \in V \tag{2.2.5}$$

(Note that the norms here are V-norms). It is symmetric *if $\mathsf{A}(f,g) = \mathsf{A}(g,f)$ for all $f,g \in V$.*

Definition 2.2.2 (Variational equivalent of a PDE)**.** *Let A be a $2k$-degree differential operator. If A is positive semi-definite, then we can define a bounded bilinear form on H^k by*

$$\mathsf{A} := \langle \cdot | A | \cdot \rangle$$

We can then search for solutions to the differential equation $Au = f$ in a subspace *$V \subseteq H^k$, on which this operator is well defined. Letting now $F = |f\rangle$, we can write the differential equation on the* variational *form*

$$\mathsf{A}(u,v) = F(v) \quad \forall v \in V \tag{2.2.6}$$

The following lemma gives sufficient conditions for solutions to Equation (2.2.6) to be well defined:

Lemma 2.2.3 (Existence and uniqueness of solutions)**.** *Let V be a subspace of $H^k(\Omega)$, $F : V \to \mathbb{R}$ be a continuous functional on V, and let $\mathsf{A} : V \times V \to \mathbb{R}$ be a bounded, symmetric, and elliptic bilinear form on V. Then Equation (2.2.6)*

$$\mathsf{A}(u,v) = F(v) \quad \forall v \in V$$

has a unique solution $u \in V$.

Proof. The conditions of A being symmetric, bounded and elliptic directly imply that A defines an inner product that is equivalent to the inner product on V. Since V is a Hilbert space, it is then also a Hilbert space under the inner product A, and we can call this space V_A.

Since V_A is a Hilbert space, the Riesz Representation Theorem states that the dual space V_A^*, consisting of the bounded linear functionals, is exactly the space of inner products $\mathsf{A}(u,\cdot) = \langle u|_A$. In particular, this implies that there exists a unique $u \in V$ such that

$$\mathsf{A}(u,v) = F(v) \quad \forall v \in V$$

QED. □

Example 2.2.4 (Poisson's Equation)**.** Consider Poisson's equation with homogeneous Dirichlet and von Neumann boundary conditions on a subdomain $\Omega \subseteq \mathbb{R}^n$:

$$\begin{aligned} -4\pi\nabla^2 u(\mathbf{x}) &= \rho(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega \\ u(\mathbf{x}) &= 0 \qquad \forall \mathbf{x} \in \partial\Omega \\ \frac{\partial}{\partial \mathbf{n}} u(\mathbf{x}) &= 0 \qquad \forall \mathbf{x} \in \partial\Omega \end{aligned} \tag{2.2.7}$$

Where $\mathbf{n}(\mathbf{x})$ is normal to the boundary surface $\partial\Omega$. We can state Equation (2.2.7) in variational form by choosing $V = \left\{ f \in H^1(\Omega) \;\middle|\; f|_{\partial\Omega} = \frac{\partial}{\partial \mathbf{n}} = 0 \right\}$ and setting

$$\begin{aligned} \mathsf{A}(u,v) &= -4\pi \int_\Omega d^n x \; \nabla u \nabla v \\ F(v) &= \int_\Omega d^n x \; \rho \; v \end{aligned} \tag{2.2.8}$$

and thus searching for a solution $u \in V$ that satisfies the standard form

$$\mathsf{A}(u, v) = F(v) \quad \forall v \in V$$

The necessity of the boundary conditions in the partial differential equation (2.2.7) is reflected in the necessity of restricting the space for the ellipticity condition in Lemma 2.2.3 to hold. Consider what happens if we set $V = H^1(\Omega)$: It can be readily seen that A is bilinear and symmetric, and the boundedness on H^1 (and thus on V) of A and F follows directly from the definition of H^1. However, A is not elliptic on H^1. This can be readily verified by considering piecewise linear functions $l(\mathbf{x})$ on Ω, which are clearly in H^1. For $l(\mathbf{x}) \neq 0$, these all have nonzero H^1-norms, but $\mathsf{A}(l, l) = 0$, whereby A cannot be elliptic: $\mathsf{A}(x, x)$ is only a pseudo-norm on H^1. This corresponds to the fact that solutions to Equation (2.2.7) are not unique: If u is any solution, then whenever $\mathsf{A}(f, f) = 0$, the sum $u + f$ is an equally good solution.

For many other examples of variational forms of partial differential equations, see for example Brenner and Carstensen [2004], Section 2.

In Section 2.3.4, we will review how more elaborate boundary conditions are handled.

2.2.1 Ritz-Galerkin methods

Let a variational problem of the form (2.2.6) be given, with solutions in the Hilbert space V. If we now let $\tilde{V}$ be a finite-dimensional subspace to V, we can pose a *discrete* approximation to the problem: Find $\tilde{u} \in \tilde{V}$ such that

$$\mathsf{A}(\tilde{u}, \tilde{v}) = F(\tilde{v}) \quad \forall \tilde{v} \in \tilde{V} \tag{2.2.9}$$

Solving Equation (2.2.9) amounts to computing the matrix representing A and the vector for F in some basis for $\tilde{V}$ and then solving the linear system. By subtracting (2.2.9) from (2.2.6), we immediately obtain

$$\mathsf{A}(u - \tilde{u}, \tilde{v}) = 0 \quad \forall \tilde{v} \in \tilde{V} \tag{2.2.10}$$

That is: the approximate solution $\tilde{u}$ is the projection of the exact solution u onto $\tilde{V}$ under the inner product A. In this sense, the approximate solution is the "best" solution in $\tilde{V}$, when the error is measured in the A-norm. Furthermore, the boundedness and ellipticity of A then immediately yield an a priori bound on the error, given in the A-norm:

$$\|u - \tilde{u}\|_{\mathsf{A}} \leq \sqrt{\frac{C_1}{C_2}} \inf_{\tilde{v} \in \tilde{V}} \|u - \tilde{v}\|_{\mathsf{A}} \tag{2.2.11}$$

which shows that the approximate solution $\tilde{u}$ is optimal up to the constant $\sqrt{\frac{C_1}{C_2}}$ in the A-norm. These bounds on the error hold without reference to anything other than this very general framework. Much tighter bounds can be found when we specialize to finite element methods. We will only touch briefly on these in Section 2.3.5, but we point the interested reader towards Brenner and Scott [2008], which contains extensive treatments of error bounds.

2.3 Finite Element discretizations of space

Finite element methods fit naturally into the framework of Sobolev spaces: They are a special class of Ritz-Galerkin function space discretization methods that yield both efficient and accurate results. The overall idea is to divide the domain Ω into a finite number of disjoint subdomains $\bigcup_{i=1}^{N_{\text{cells}}} c_i = \Omega$, which we shall call *cells*.

The finite basis $\{ \mathfrak{e}_i \mid 1 \leq i \leq N_{\text{dof}} \}$ has two essential properties:

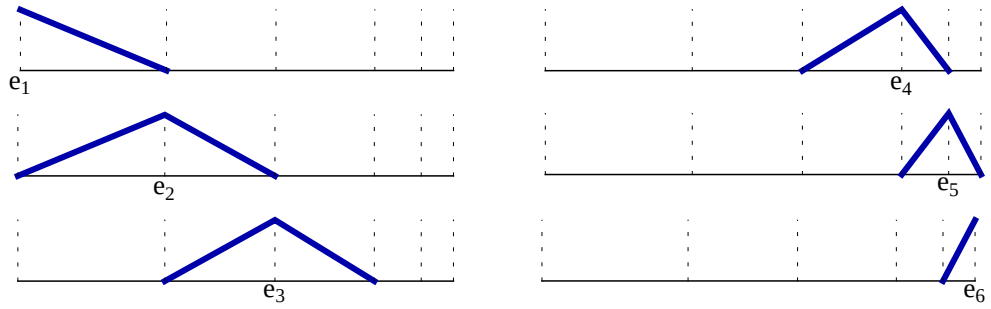


Figure 2.1: Example of a piecewise linear finite element basis in one dimension. The $\mathfrak{e}_i$ are the global finite element basis functions.

1. Every basis function $\mathfrak{e}_i$ is a polynomial when restricted to any cell c, i.e. it is a piecewise polynomial, and the non-differentiable points are on cell boundaries.
2. Each basis function $\mathfrak{e}_i$ has support only on a few neighbouring cells, giving rise to extremely sparse systems of equations.

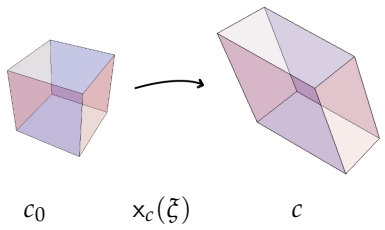
We will often refer to these basis functions as the *global* basis functions, in order not to confuse them with the per-cell basis functions (or *shape functions*) described in the next section. Further, we will sometimes refer to them as DOFs ("Degrees of Freedom"), as is done by Bangerth et al.. This is so as not to cause confusion between the finite element basis functions with the basis functions used in the quantum theoretical framework.

Because basis functions are piecewise polynomials, any function in their span is infinitely differentiable under the weak derivative. A finite element space is hence always a subspace of $H^\infty(\Omega)$, and thus by transitivity is a subspace of all of the entire sequence $H^1(\Omega) \supseteq H^2(\Omega) \supseteq \cdots$ of Sobolev Hilbert spaces. In particular, any finite element space is itself a Sobolev space and a Hilbert space. Furthermore, the set of finite element spaces is dense in H^∞, which in turn is dense in every H^k, similar to how the space of usual smooth functions C^∞ is dense in C^k for all k. Because of this, we can obtain arbitrarily good approximations by using sufficiently finely divided finite element spaces or sufficient polynomial degrees. Figure 2.1 shows an example of a piecewise linear finite element discretization of functions on a line segment. The basis functions $\{\mathfrak{e}_1, \ldots, \mathfrak{e}_6\}$ span a six-dimensional subspace to $H^\infty([a;b])$.

2.3.1 The unit cell and local shape functions

A third important property of finite element methods is the concept of the *unit cell*. This is the source of much of the computational ease of finite element methods, yielding not only very sparse but also structured systems of equations.

The unit cell is a simple reference cell c_0 from which every other cell c are derived through a per-cell (usually linear) coordinate transformation x_c:



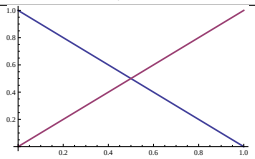
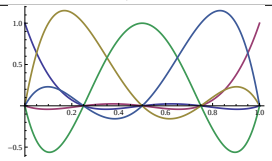
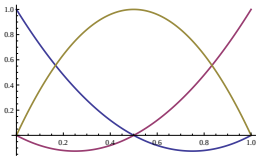
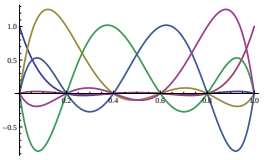
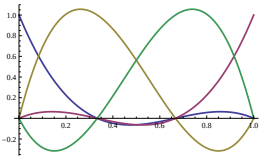
FE order	e_i's	FE order	e_i's
1		4	
2		5	
3		⋮	

Table 2.1: Shape functions e_i for one-dimensional finite elements on the unit line

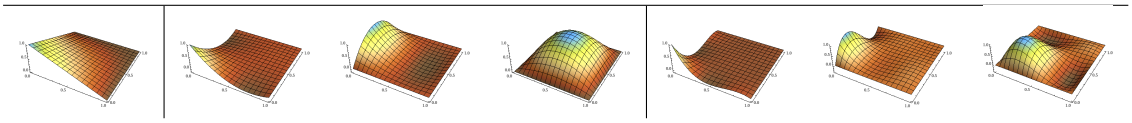
1	2	3
		

Table 2.2: Two-dimensional local shape functions e_i.

Note that *unit cell* is a slight misnomer, since is not in fact always of unit measure. However, we follow the terminology used by Bangerth et al.. In the work for this thesis, we have used *tensor-product elements*. What this means, is that n-dimensional basis functions are defined as simple tensor products of one-dimensional basis functions:

$$\mathfrak{e}_{i_1,\ldots,i_n}(x_1,\ldots,x_n) = \mathfrak{e}_{i_1}(x_1)\cdots\mathfrak{e}_{i_n}(x_n)$$

which naturally leads to the unit cell being the tensor product of a line segment with itself n times, yielding an n-dimensional cube. This makes for the simplest theoretical framework. Many other choices are possible; the most commonly used finite element schemes use tetrahedral cells. However, little in this thesis depends on the particular choice.

On the unit element are defined a small fixed number of local basis functions $\{ e_i \mid 1 \leq e_i \leq n_{\text{dof}} \}$, which we call *shape functions*. These form a basis of the polynomial space that is spanned on the cell, and define the piecewise polynomial basis for the whole finite element space: The polynomial basis for any given cell c is simply the set of shape functions transformed through x_c^{-1}: $\{ e_i \circ \mathsf{x}_c^{-1} \mid 1 \leq e_i \leq n_{\text{dof}} \}$. Consider, for example, the case of one-dimensional piecewise linear spaces as we saw in Figure 2.1, and let the unit cell $c_0 = [0;1]$. The shape functions from which the basis functions are constructed are then $e_1(x) = x$ and $e_2(x) = 1 - x$. The cell transforms x_c from the unit cell coordinate system to the global coordinate system are here simply a scaling followed by a translation.

Table 2.1 shows the shape functions used in this work for polynomial degrees 1–5. Shape functions in higher dimensional spaces are then constructed as simple tensor products of the one-dimensional shape functions. Figure 2.2 shows a selection of two-dimensional shape functions for polynomial degrees 1–3.

To each global basis function e_i is associated a *nodal position* n_i such that $\mathfrak{e}_i(n_i) = 1$ and $\mathfrak{e}_i(n_j) = 0$ when $i \neq j$. The nodal positions thus depend on the polynomial basis, and $\mathfrak{e}_i(n_j) = \delta_{ij}$, yielding

the relation

$$\tilde{f} = \sum_{i=1}^{N_{\text{dof}}} a_i \mathfrak{e}_i \implies a_i = \tilde{f}(n_i) \tag{2.3.1}$$

That is, the coefficients and the values at nodal positions are the same. A global basis functions $\mathfrak{e}_i$ has support only on cells adjacent to its nodal position, and is zero everywhere else. This gives a fixed and very regular structure to the systems of equations that can be exploited.

2.3.2 Integration on finite element spaces

One- and two-function integrals of finite-element functions are incredibly simple $\mathcal{O}(N_{\text{dof}})$-operations due to sparsity. Let $\tilde{f} = \sum_{i=1}^{N_{\text{dof}}} f_i \mathfrak{e}_i$ and $\tilde{g} = \sum_{i=1}^{N_{\text{dof}}} g_i \mathfrak{e}_i$ be two finite element functions. Then we trivially have

$$\begin{aligned} \int_\Omega d^n x\, \tilde{f}(\mathbf{x}) &= \sum_{i=1}^{N_{\text{dof}}} f_i \int_\Omega d^n x\, \mathfrak{e}_i(\mathbf{x}) \\ \int_\Omega d^n x\, \tilde{f}(\mathbf{x})\tilde{g}(\mathbf{x}) &= \sum_{i,j=1}^{N_{\text{dof}}} f_i g_j \int_\Omega d^n x\, \mathfrak{e}_i(\mathbf{x})\mathfrak{e}_j(\mathbf{x}) \end{aligned} \tag{2.3.2}$$

If we define the vector $\mathbf{m}$ and the (very sparse) matrix $\mathbf{M}$ by $m_i = \int_\Omega d^n x\, \mathfrak{e}_i(\mathbf{x})$ and $M_{ij} = \int_\Omega d^n x\, \mathfrak{e}_i(\mathbf{x})\mathfrak{e}_j(\mathbf{x})$, the integrals are simply a dot product in the one case and a sparse matrix-vector product and dot product in the second case:

$$\begin{aligned} \int_\Omega d^n x\, \tilde{f}(\mathbf{x}) &= \mathbf{m}^T \mathbf{f} \\ \int_\Omega d^n x\, \tilde{f}(\mathbf{x})\tilde{g}(\mathbf{x}) &= \mathbf{f}^T \mathbf{M} \mathbf{g} \end{aligned} \tag{2.3.3}$$

The integrals are exact, since we can precompute $\mathbf{m}$ and $\mathbf{M}$ exactly. Similarly, we can write

$$\int_\Omega d^n x\, \nabla\tilde{f}(\mathbf{x}) \cdot \nabla\tilde{g} = \sum_{i,j=1}^{N_{\text{dof}}} f_i g_j \int_\Omega d^n x\, \nabla\mathfrak{e}_i(\mathbf{x}) \cdot \nabla\mathfrak{e}_j(\mathbf{x}) = \mathbf{f}^T \mathbf{T} \mathbf{g} \tag{2.3.4}$$

and so forth.

We now turn to the problem of integrating arbitrary functions that are given point-wise rather than as expansions in finite element basis functions. This also adresses the problem of computing $\mathbf{m}$, $\mathbf{M}$, $\mathbf{T}$, et cetera. This is achieved by *quadrature* and transformation of the unit cell.

We break down the integral over Ω by splitting it into a sum of the contributions from each cell. Consider the integral of an arbitrary integrable function f on any given cell c. We can rewrite this as an integral on the unit cell c_0 as

$$\int_c d^n x\, f(\mathbf{x}) = \int_{c_0} d^n \xi\, f(\mathsf{x}_c(\xi))|J_c(\xi)| \tag{2.3.5}$$

where J_c is the Jacobian of the mapping x_c. We now introduce a quadrature in order to evaluate the integral on the unit cell numerically. A quadrature formula is a tuple of points ξ_q on the unit cell and a tuple of weights w_q, such that

$$\int_{c_0} d^n \xi\, f(\xi) \approx \sum_{q=1}^{n_q} w_q\, f(\xi_q) \tag{2.3.6}$$

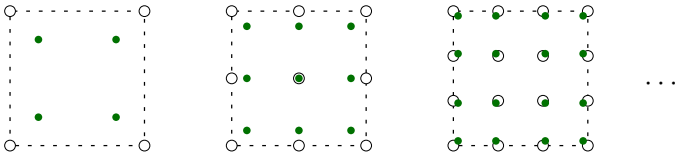


Figure 2.2: Unit squares for two-dimensional tensor product elements. The white circles show the nodal positions, at which the shape functions $e_i(\mathbf{x}) = 1$, and the green dots are the Gaussian quadrature points. The shape functions are Legendre polynomials and are shown for one dimension in Table 2.1.

We have used the Gaussian quadrature for this work. An n-point (per dimension) Gaussian quadrature is constructed in such a way as to yield exact integrals up to polynomials of degree $(2n-1)$ (per dimension). That is,

$$\int_{c_0} d^n\xi\, f(\xi) = \sum_{q=1}^{n_q} w_q\, f(\xi_q)$$

whenever f restricted to c_0 is a polynomial of degree no more than $2n_q - 1$. Figure 2.2 shows the unit cells for two-dimensional tensor product elements, together with the nodal positions for Legendre polynomial shape functions of degree 1–3, and Gaussian quadrature points appropriate for integrating products of the shape functions.

Now we can evaluate integrals on any cell. Before proceeding, let us introduce a notational shortcut: When the cell c is fixed, let $i = loc_c(I)$. That is, e_i is the shape function such that $e_i \circ \mathsf{x}_c$ is equal to the restriction of $\mathfrak{e}_I$ to c. We will use upper case for the global indices and implicitly assume that the lower case letter is the corresponding local index — and vice versa.

Now, if $\mathfrak{e}_I$ has support on the cell c, then the cell contribution m_i^c to m_I is

$$\begin{aligned} m_i^c = \int_c d^n x\, \mathfrak{e}_i(\mathbf{x}) &= \int_c d^n x\, e_i(\mathsf{x}_c^{-1}(\mathbf{x})) \\ &= \int_{c_0} d^n\xi\, |J_c(\xi)|\, e_i(\xi) \\ &= \sum_{q=1}^{n_q} \underbrace{w_q |J_c(\xi_q)|}_{w_{cq}}\, e_i(\xi_q) \\ &= \sum_{q=1}^{n_q} w_{cq}\, e_{iq} \end{aligned} \tag{2.3.7}$$

Similarly,

$$\begin{aligned} M_{ij}^c = \int_c d^n x\, \mathfrak{e}_i(\mathbf{x})\mathfrak{e}_j(\mathbf{x}) &= \int_{c_0} d^n\xi\, |J_c(\xi)|\, e_{i)}(\xi) e_j(\xi) \\ &= \sum_{q=1}^{n_q} w_{cq}\, e_{iq}\, e_{jq} \end{aligned} \tag{2.3.8}$$

when n_q is large enough that products $\mathfrak{e}_I \cdot \mathfrak{e}_J$ are still of polynomial degree at most $2n_q - 1$,

$$\int_c d^n x\, \mathfrak{e}_I(\mathbf{x}) f(\mathbf{x}) \mathfrak{e}_J(\mathbf{x}) = \sum_{q=1}^{n_q} w_{cq}\, e_{iq}\, f(\mathsf{x}_c(\xi_q))\, e_{jq} \tag{2.3.9}$$

Symbol	Description
$\tilde{V}$	Our finite-element space.
DOF	(Short for "Degree of freedom"): The (global) finite element basis functions.
N_{cells}	Total number of cells.
N_{dof}	Total number of global basis functions (dofs).
n_{dof}	Number of local basis functions (shape functions) per cell.
N_q	Total number of quadrature points.
n_q	Number of quadrature points per cell.
$\mathfrak{e}_i$	The i'th *global* basis function (DOF).
e_j	The j'th *local* shape function.
$glob_c(j)$	The global DOF index corresponding to the j'th shape function on cell c.
$loc_c(i)$	The local shape function index corresponding to the i'th global DOF on cell c.
w_q	The weight associated with the q'th quadrature point on the unit cell.
x_q	The q'th global quadrature point.
ξ_q	The q'th local quadrature point (in the unit cell coordinates).
x_{cq}	The q'th quadrature point of cell c in global coordinates.
$\mathbf{M}$	The (local) overlap matrix $\mathbf{M}_{ij} = \int_c dx\, e_i(x) e_j(x)$.

Table 2.3: Overview of the terminology and corresponding symbols used in this chapter and Chapter 5.

and so forth. Whenever the number of quadrature points matches the polynomial degree of the integrand on the cell, all the integrals are exact. If the polynomial degree is too large or indeed if the integrand is superpolynomial, they are only approximate.

2.3.3 Solving finite element problems

Example 2.3.1 (Revisiting Poisson's Equation)**.** Recall the variational form of Poisson's equation in n dimensions with a dielectric function $\epsilon(\mathbf{x})$:

$$\begin{aligned} \mathsf{A}(u,v) &= -4\pi \int_\Omega d^n x\, \epsilon(\mathbf{x}) \nabla u \cdot \nabla v \\ F(v) &= \int_\Omega d^n x\, \rho(\mathbf{x}) v(\mathbf{x}) \end{aligned} \tag{2.3.10}$$

Let there now be given a finite element discretization of $V \subseteq H^1(\Omega)$ that preserves the discontinuities of $\epsilon(\mathbf{x})$, i.e. all its discontinuities lie along cell faces. We can then write

$$\begin{aligned} A_{ij} &= -4\pi \int_\Omega d^n x\, \epsilon(\mathbf{x}) \nabla \mathfrak{e}_i(\mathbf{x}) \cdot \nabla \mathfrak{e}_j(\mathbf{x}) \\ &= -4\pi \sum_{c \in Cells} \int_c d^n x\, \epsilon(\mathbf{x}) \nabla \mathfrak{e}_i(\mathbf{x}) \cdot \nabla \mathfrak{e}_j(\mathbf{x}) \\ &= \sum_c \sum_{q=1}^{n_q} \epsilon_c(\xi_q) \nabla e_{loc_c(i)}(\xi_q) \cdot \nabla e_{loc_c(j)}(\xi_q) w_{cq} \end{aligned} \tag{2.3.11}$$

where in the last line we sum over all cells on which both $\mathfrak{e}_i$ and $\mathfrak{e}_j$ have nonzero values. This is a somewhat cumbersome angle from which to attack the problem. A less convoluted approach that exploits the sparsity of the problem, is to construct the full matrix $\mathbf{A}$ cell by cell. For each cell c, we construct a small, dense $n_{\text{dof}} \times n_{\text{dof}}$ matrix indexed by the local shape functions:

$$A^c_{ij} = \sum_{q=1}^{n_q} \epsilon_c(\xi_q) \nabla e_i(\xi_q) \cdot \nabla e_j(\xi_q) w_{cq} = \sum_{q=1}^{n_q} \epsilon_c(\xi_q) w_{cq} T^q_{ij} \tag{2.3.12}$$

In the rightmost expression, we have used the fact that the quantity $T_{ij}^q = \nabla e_i(\xi_q) \cdot \nabla e_j(\xi_q)$ is a property of the unit cell and so never changes. It is a small $n_q \times n_{\text{dof}}^2$ matrix and can be precalculated. The c-contribution to the vector $\mathbf{F}$ is

$$F_i^c = \sum_{q=1}^{n_q} \rho(x_q) e_i(\xi_q) w_{cq} \tag{2.3.13}$$

Constructing the sparse linear system then amounts to iterating through the cells, for each cell c constructing the (dense) cell-matrix $\mathbf{A}^c$ and vector $\mathbf{F}^c$, then adding A_{ij}^c to A_{IJ} and F_i^c to F_I, where $(I, J) = (glob_c(i), glob_c(j))$ are the global DOF indices.

Once the global matrix $\mathbf{A}$ and right hand side vector $\mathbf{F}$ are constructed, the sparse linear system

$$\mathbf{Au} = \mathbf{F}$$

is solved to obtain the vector $\mathbf{u}$ such that $u(\mathbf{x}) = \sum_i u_i \mathfrak{e}_i(\mathbf{x})$ is the solution to the variational problem (2.3.10) on the finite element space.

While we have not shown the general case, but have concentrated on Poisson's equation, it should hopefully not be difficult for the reader to derive from the example how to do other variational problems.

2.3.4 Adding boundary conditions

For the sake of keeping the introductory example simple, we have avoided the talk of boundary conditions in Example 2.3.1. However, it is not difficult to alter the example in order to include boundary conditions:

Example 2.3.2 (Poisson's equation complex boundary conditions)**.** Consider Example 2.3.1, but with the addition of (possibly inhomogeneous) Dirichlet boundary conditions and (possibly inhomogeneous) von Neumann boundary conditions:

$$\begin{aligned} -4\pi\epsilon(\mathbf{x})\nabla^2 V(\mathbf{x}) &= \rho(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega \\ V(\mathbf{x}) &= f(\mathbf{x}) \quad \forall \mathbf{x} \in \Gamma \subseteq \Omega \\ \frac{\partial V(\mathbf{x})}{\partial \mathbf{n}} &= g(\mathbf{x}) \quad \forall \mathbf{x} \in \partial\Omega \setminus \Gamma \end{aligned} \tag{2.3.14}$$

This is the most general case that we will need for this thesis. Notice that the Dirichlet conditions are not restricted to the boundary $\partial\Omega$ of the domain: Γ may be any subset of Ω, as indeed we will find to be necessary later. However, we *do* restrict the von Neumann conditions to the boundary. We now assume that we are given a finite element discretization of Ω that respects both Γ and ϵ, i.e. all discontinuities of $\epsilon(\mathbf{x})$ as well as the boundary of Γ coincide with cell faces.

The two types of boundary conditions can be described as transformations of the unrestricted problem defined by Equations (2.3.12) and (2.3.13). They can be applied independently, and we will treat them separately.

The Dirichlet constraints are applied as follows: We call the set of DOFs that have support in Γ the *fixed* DOFs, and write this set as $\text{fixed}(\Gamma)$. The fixed variables do do not need to be solved for in the linear system, but their values should flow to their neighbours as usual. For the matrix $\mathbf{A}$, this corresponds to zeroing out the corresponding rows , which describe data flow *into* a variable, while leaving untouched the columns, which describe data flow *out of* a variable. The diagonal

element is also left in place. In the right hand side vector $\mathbf{F}$, we set $F_i = f_i / A_{ii}$ for fixed indices i, ensuring that the solution will have $u_i = f_i$:

$$\begin{aligned} \text{Dirichlet}(\Gamma, \mathbf{A})_{ij} &= \begin{cases} A_{ij} & \text{for } i \notin \text{fixed}(\Gamma) \text{ or } i = j \\ 0 & \text{for } i \in \text{fixed}(\Gamma) \text{ and } i \neq j \end{cases} \\ \text{Dirichlet}(\Gamma, \mathbf{f}, \mathbf{F})_i &= \begin{cases} F_i & \text{for } i \notin \text{fixed}(\Gamma) \\ f_i / A_{ii} & \text{for } i \in \text{fixed}(\Gamma) \end{cases} \end{aligned} \tag{2.3.15}$$

The von Neumann boundary condition affects only the right hand side in Equation (2.3.10), for which a surface integral is added:

$$F(v) = \int_\Omega d^n x \, \rho(\mathbf{x}) v(\mathbf{x}) + \int_{\partial\Omega \setminus \Gamma} d^{n-1} s \, \epsilon(\mathbf{s}) g(\mathbf{s}) \tag{2.3.16}$$

For internal cells or surface cells not on $\partial\Omega \setminus \Gamma$, Equation (2.3.13) is unchanged; for shape functions on $\partial\Omega \setminus \Gamma$, the surface integral is added to F_i^c.

Note that the homogeneous von Neumann boundary condition corresponds to the special case $g(\mathbf{s}) = 0$ in Equation (2.3.16), i.e. if nothing is specified, the solution will have homogeneous von Neumann boundary conditions. Because of this, homogeneous von Neumann conditions are often called *natural* boundary conditions.

2.3.5 Error estimates

A priori error estimates are are properties of a certain finite element space discretization, while methods for a posteriori error estimates use a given finite element solution to derive stricter bounds on the error relative to the exact solution. I will not give a detailed treatment of error estimates here, but only cite the most important inequalities. There exists extensive literature on the subject; an excellent treatment is given in Brenner and Scott [2008], Chapters 2, 4, 8, and 9. A shorter review is given in Brenner and Carstensen [2004], Section 4.

Define the *mesh-size function* $h(\mathbf{x})$ as the piecewise constant function

$$h(\mathbf{x}) := \max_{\mathbf{x} \in c} \text{diameter}(c) \tag{2.3.17}$$

that is, the mesh-size at $\mathbf{x}$ is the diameter of the largest cell that $\mathbf{x}$ touches.

We then consider Poisson's Equation with a dielectric function $\epsilon(\mathbf{x})$, i.e.

$$\mathsf{A}(u, v) = \int_\Omega d^3 x \, \epsilon(\mathbf{x}) \nabla u \cdot \nabla v$$

and we assume that the discontinuities in $\epsilon(\mathbf{x})$ coincide with faces in the mesh.

The most important a priori estimate for our purposes is the following:

Theorem 2.3.3 (Brenner and Scott [2008], T9.1.10)**.** *There exists a constant $\kappa > 0$ such that if the mesh size variation, $\| |\nabla h| \|_{L^\infty(\Omega)} < \kappa$, then*

$$\|u - u_h\|_{L^2(\Omega)} \leq C \left\| \sqrt{\epsilon} h^2 |\nabla^2 u| \right\|_{L^2(\Omega)}$$

where $C > 0$ is a constant that is independent of the mesh.

One noteworthy implication of this is, that if the mesh is sufficiently regular, i.e. if the maximal variation of the mesh-size function is small enough, then the finite element solution u_h to Poisson's equation has error relative to the real solution u bounded by $\mathcal{O}(\|h^2\|)$. In particular, for a given cell, the cell's contribution to the upper bound to the error is quadratic in the cell diameter and linear in the amount of curvature $\||\nabla^2 u|_c|\|$ contained in the cell. The value of κ depends of course on the finite element degree and other specifics of the elements that are used.

In the present work, we use finite element spaces with hexagonal elements. Recursive refinement is performed only by way of regular subdivision, i.e. a line is evenly split in two, a rectangle in four, and a hexagon in eight equally sized pieces, and neighbouring cells may only differ by one level of subdivision. Thus, if the original coarse mesh is sufficiently regular, the ratio between the diameters of any two adjacent cells c_1 and c_2 in the refined mesh is bounded by one half and two:

$$\frac{1}{2} \leq \frac{\operatorname{diam}(c_1)}{\operatorname{diam}(c_2)} \leq 2 \tag{2.3.18}$$

We have not yet investigated a posteriori error estimates for the methods that have been developed; This is relegated to future work.

CHAPTER 3

A brief review of density functional theory

3.1 The Schrödinger equation and a bit of notation

The fundamental physical properties of matter are governed by a wave equation, the Schrödinger Equation in the non-relativistic approximation, and the Dirac Equation when relativity is included. For a fixed number N of particles, i.e. when disallowing creation and annihilation, the non-relativistic, time independent Schrödinger equation can be written in atomic units as

$$[-\tfrac{1}{2}\Delta + \mathsf{W} + \mathsf{V} - E]\,\Psi = 0, \quad \text{where} \quad \Delta := \sum_{j=1}^{N} \frac{1}{m_j}\nabla_j^2 \tag{3.1.1}$$

When the energy E and wave function Ψ are unknown, this is the eigensystem

$$\mathsf{H}\Psi = E\Psi \tag{3.1.2}$$

where $\mathsf{H} := -\frac{1}{2}\Delta + \mathsf{W} + \mathsf{V}$ is called the *Hamiltonian operator*. In the above, ∇_j^2 works on the j^{th} particle (of mass m_j). W is called the *internal* potential or the *interaction* potential, and it is the potential arising from the mutual forces exerted between the particles. V acts separately on individual particles and is often called the *external* potential.

Example 3.1.1 (Molecule)**.** The bound states of a molecule with N_e electrons and N_I nuclei with charges $Z_1, \ldots, Z_{N_I}$ are described by Equation (3.1.1) with

$$\mathsf{W} = \sum_{i=1}^{N_e}\sum_{j=i+1}^{N_e} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} + \sum_{I=1}^{N_I}\sum_{J=I+1}^{N_I} \frac{Z_I Z_J}{|\mathbf{X}_I - \mathbf{X}_J|} - \sum_{i=1}^{N_e}\sum_{J=1}^{N_I} \frac{Z_I}{|\mathbf{X}_I - \mathbf{x}_i|} \tag{3.1.3}$$

and $\mathsf{V} = V^{\text{ext}}(\mathbf{x})$ is whatever external potential in which the molecule is situated.

Example 3.1.2 (Molecule, Born-Oppenheimer approximation)**.** Because the nuclei are thousands of times heavier than electrons, and acceleration is force over mass, we can assume to a good approximation that electronic motion is instantaneous compared to the nuclear motion. This gives rise to one of the most commonly used schemes in quantum theory: The Born Oppenheimer approximation. In this approximation, the electronic system is solved assuming fixed nuclei. This is done for a range of nuclear positions. The electronic solutions yield a potential surface in which the nuclei can move, and the Schrödinger equation is solved for only the nuclei, with the precomputed electronic potential acting as an external potential. Thus, for the electronic Schrödinger

equation,

$$\mathsf{W}_e = \sum_{i=1}^{N_e} \sum_{j=i+1}^{N_e} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} \quad \text{and} \quad \mathsf{V}_e = -\sum_{i=1}^{N_e} \sum_{J=1}^{N_I} \frac{Z_I}{|\mathbf{X}_I - \mathbf{x}_i|} \tag{3.1.4}$$

and for the nuclei:

$$\mathsf{W}_I = \sum_{I=1}^{N_I} \sum_{J=I+1}^{N_I} \frac{Z_I Z_J}{|\mathbf{X}_I - \mathbf{X}_J|} \quad \text{and} \quad \mathsf{V}_e = -\sum_{J=1}^{N_I} Z_I V^{el}[\mathbf{X}_1, \ldots, \mathbf{X}_{N_I}](\mathbf{X}_I) \tag{3.1.5}$$

It is common, as well, to treat the heavy nuclei as classical particles moving in the computed electronic potential surface. However, we will not consider this side of the problem: The methods developed in this thesis deal only with the solution of electronic systems, which in general is the most computationally intensive part of a quantum chemical calculation. We will assume that a system consists of N electrons moving in the external potential

$$V(\mathbf{x}) = \sum_{i=1}^{N} v(\mathbf{x}_i) = -\sum_{i=1}^{N} \sum_{I=1}^{N_I} \frac{1}{|\mathbf{X}_I - \mathbf{x}_i|} + \sum_{i=1}^{N} v^{\text{ext}}(\mathbf{x}_i) \tag{3.1.6}$$

where $v(\mathbf{x}_i)$ acts on the i^{th} electron.

$$W(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{N} \sum_{j=i+1}^{N} w(\mathbf{x}_i, \mathbf{x}'_j) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{|\mathbf{x}_i - \mathbf{x}'_j|} \tag{3.1.7}$$

We will in general use capital letters for N-particle potentials, and lower case letters for their constituent one- and two-particle potentials.

3.2 The need for DFT

Only the tiniest systems have known solutions to the Schrödinger Equation, and approximations must be sought. However, even solving the many-body equation approximately is a daunting task. The N-electron wave equation is a partial differential equation with $3N$ free variables, and no separation of variables is possible due to electron interaction: The wave function is a true $3N$-dimensional function. Commonly, Hilbert space methods are used, and solutions are sought in the antisymmetrized product space

$$\mathfrak{H}_N = \bigwedge^N H^1(\mathbb{R}^3; \mathbb{K})$$

where the function values are $\mathbb{K} = \mathbb{C}^2$ if spin is included, and $\mathbb{K} = \mathbb{C}$ otherwise. If the one-electron space $H^1(\mathbb{R}^3)$ is approximated with a finite dimensional space $\tilde{\mathfrak{h}}$ with dimension M, then the full approximate N-electron space $\tilde{\mathfrak{H}}_N$ has dimension

$$\dim\left(\tilde{\mathfrak{H}}_N\right) = \binom{M}{N}$$

That is, the dimension of the N-electron space grows combinatorially in the dimension of the one-electron approximation. A brute force approach is thus clearly not feasible, and much of computational quantum theory is dedicated to sidestepping this barrier.

The present chapter covers the basics of *density functional theory*, the aim of which is to compute quantum theoretical properties of matter in a way that avoids solving the many-body problem. Instead, ground state properties are computed directly from the ground state density, a simple function of space and spin, without the need of ever obtaining the full N-electron wave function.

That this is in fact possible is a surprising and unintuitive result. It was originally proposed by Pierre Hohenberg, Walter Kohn, Lu Jeu Sham during the mid 1960s, but the theory did not spring into life fully formed: Roughly twenty years passed after the seminal paper by Hohenberg and Kohn, before the Hohenberg-Kohn theorems were proved in a mathematically sound way.

3.3 The Hohenberg-Kohn theorems

The Hohenberg-Kohn Theorems, proposed in Hohenberg and Kohn [1964], form the basis for all density functional theory. They state a quite bold and unintuitive claim: that the (averaged) electron density[1]

$$\rho[\Psi_0](\mathbf{x}) := \int d^3x_2 \cdots \int d^3x_N |\Psi(\mathbf{x}, \mathbf{x}_2, \ldots, \mathbf{x}_N)|^2 \tag{3.3.1}$$

of a ground state Ψ_0 contains sufficient information to determine the N-electron Hamiltonian from which it arose, and thus, theoretically, every property of the physical system. The theorems were originally proposed only for non-degenerate ground states, in which case the first theorem asserts the existence of a map $\Psi[\rho]$ that reconstructs a ground state from its density. This in turn yields, for any observable A, a *functional*

$$A[\rho] := \langle \Psi[\rho] | \mathsf{A} | \Psi[\rho] \rangle \tag{3.3.2}$$

that maps the density to the corresponding expectation value of A. These are called *density functionals*, from which density functional theory takes its name. If we allow degenerate ground states, the situation is slightly muddled. In this case, we must choose which of the N-electron wave functions that produce ρ to use. We will then in general let

$$A[\rho] := \min_{\Psi \mapsto \rho} \langle \Psi | \mathsf{A} | \Psi \rangle \tag{3.3.3}$$

and most of the theory carries through unscathed.

We state the first Hohenberg-Kohn theorem in the following form:

Theorem 3.3.1 (Hohenberg-Kohn Theorem I)**.** *If two N-electron systems, with well-behaved[2] external potentials $v_1(\mathbf{x})$ and $v_2(\mathbf{x})$ respectively, have the same ground state density $\rho_0(\mathbf{x})$, then $v_1(\mathbf{x}) = v_2(\mathbf{x}) + C$ for some constant C.*

We will only state the outline of the proof here, since it can be found many places in the literature (although many of the given "proofs" lack crucial steps). A valid proof can be found e.g. in van Leeuwen [1994], Section 2.5 and Chapter 4. The theorem rests on the *unique continuation theorem*[3] for the Schrödinger equation, an implication of which is that N-electron Hamiltonians differing by more than a constant have different ground states. Thus, conversely, the ground states uniquely determine the entire Hamiltonian operator up to a constant. It is easy to show that if Ψ_1 and Ψ_2 are N-electron ground states with the same density $\rho_0(\mathbf{x})$, then Ψ_1 and Ψ_2 are both ground states of the same Hamiltonian H. But since the ground states *uniquely* determine the Hamiltonian, due to the unique continuation theorem, this implies that the ground state *density* also uniquely determines the Hamiltonian (up to a constant).

[1] For the sake of simplicity, we leave out spin coordinates. If spin is included, spin-coordinates for electron 2 through N are also integrated away.

[2] The space of one-electron potentials for which this theorem always holds is

$$L^{3/2}(\mathbb{R}^3) + L^{\infty}(\mathbb{R}^3) = \left\{ v = u + w \;\middle|\; u \in L^{3/2}(\mathbb{R}^3) \;\wedge\; v \in L^{\infty}(\mathbb{R}^3) \right\}$$

which includes, for example, Coulombic and common electromagnetic potentials. See van Leeuwen [1994], Chapter 4.

[3] Valid for potentials of the form given in Footnote 2.

Corollary 3.3.2 (Hohenberg-Kohn Corollary I)**.** *The ground state electron density $\rho_0(\mathbf{x})$ uniquely determines the N-electron Hamiltonian* H*, and consequently all wave functions, both for ground- and excited states.*

Theoretically, we can thus determine all quantum theoretical properties of an N-electron system from the ground state density $\rho_0(\mathbf{x})$. However, the proofs are not constructive: The Hohenberg-Kohn theorem merely shows that there *exists* an inverse function $\rho_0 \mapsto \mathsf{H}$, but doesn't tell us how to compute it. Nor are we told how to compute *any* property of the wave function. Indeed, there is no guarantee that this function is even computable. Even so, we do get some help from this. For any *N-representable* density, i.e. a density constructible from at least one N-electron wave function, we can define the Levy-Lieb functional

$$F[\rho] := \min_{\Psi \mapsto \rho} \langle \Psi | -\tfrac{1}{2}\Delta + \mathsf{W} | \Psi \rangle \tag{3.3.4}$$

where W is the interelectron repulsion operator. The Levy-Lieb functional is universal: It depends only on the number of electrons N, and is independent of the external potential $V(\mathbf{x})$. The second Hohenberg-Kohn theorem can then be stated as follows:

Theorem 3.3.3 (Hohenberg-Kohn Theorem II)**.** *Let the number of electrons N and the one-electron potential $v(\mathbf{x})$ be given. The energy functional defined by*

$$E[\rho] := F[\rho] + \int d^3x \; v(\mathbf{x})\rho(\mathbf{x})$$

obtains its minimum for the exact ground state density $\rho_0(\mathbf{x})$ of the Hamiltonian

$$\mathsf{H} = -\tfrac{1}{2}\Delta + \mathsf{W} + \sum_i v(\mathbf{x}_i)$$

and its minimal value is the exact ground state energy.

Note that in the integral $\int d^3x \; \rho(\mathbf{x}) v[\rho](\mathbf{x})$, the one-electron potential is used. This can intuitively be understood from a argument of dimensions: integrating weighted by $\rho(\mathbf{x})$ is analogous to summing over the electrons and yields the N-electron energy.

Theorem 3.3.3 ensures that we can find the ground state density and its energy by minimizing $E[\rho]$. While we do not have an exact and efficient computational method for $F[\rho]$, we have many ways of approximating it. Because it is universal, a great deal of effort can be expended on finding good and efficient approximations. The only system dependent part, $\int d^3x \; v(\mathbf{x})\rho(\mathbf{x})$, is simply a three dimensional problem and is easy to calculate. However, $F[\rho]$ as written in Equation (3.3.4) is still posed as a $3N$-dimensional problem, but there exist a number of ways to construct approximations to $F[\rho]$ directly from the density. The most widely used of these, and the one we shall employ, is the Kohn-Sham approach. We will review this method briefly in the next section.

3.4 Kohn-Sham theory

The *Kohn-Sham* approach to density functional theory was introduced in Kohn and Sham [1965], the year after Hohenberg and Kohn's seminal paper Hohenberg and Kohn [1964]. In the Kohn-Sham approach, we attempt to replace the many-body correlated system with a non-interacting system that yields the same ground state solutions. That is: Given the one-electron potential $V(\mathbf{x})$, we wish to find an *effective potential* $V^{KS}(\mathbf{x}) = \sum_i v^{KS}(\mathbf{x}_i)$ such that the ground state density for the non-interacting Hamiltonian

$$\mathsf{H}_0 = -\tfrac{1}{2}\Delta + V^{KS}(\mathbf{x})$$

is the same as the ground state density $\rho(\mathbf{x})$ for the full Hamiltonian

$$\mathsf{H} = -\tfrac{1}{2}\Delta + \mathsf{W} + V(\mathbf{x})$$

The first Hohenberg-Kohn theorem ensures that $V^{KS}(\mathbf{x})$, if it exists, is unique. However, it is an open question whether an appropriate potential $V^{KS}(\mathbf{x})$ exists for any $V(\mathbf{x})$, i.e. whether any ground state density of an interacting system is also a ground state density of a non-interacting system. This must be taken as an ansatz:

Proposition 3.4.1 (Kohn-Sham ansatz)**.** *If* $\mathsf{H} = -\frac{1}{2}\Delta + \mathsf{W} + V(\mathbf{x})$ *is a Hamiltonian operator that includes electron interaction, and* $\rho(\mathbf{x})$ *is a ground state density for* H*, then there exists a non-interacting Hamiltonian* $\mathsf{H}_0[\rho] = -\frac{1}{2}\Delta + V^{KS}[\rho](\mathbf{x})$ *such that* $\rho(\mathbf{x})$ *is also a ground state density for* H_0*.*

Note also that the spectrum, as well as even the ground state wave function, are different for the interacting and the non-interacting system: Only the ground state density can be assumed to be the same. Indeed, for a long time, the Kohn-Sham spectrum was believed to have no physical meaning at all. While a number of studies have had some success in finding some level of physical interpretation of the Kohn-Sham spectrum (see for example Janak [1978], Politzer and Abu-Awwad [1998], Luo et al. [2006]), the exact meaning of the Kohn-Sham eigenvalues and orbitals has not yet been fully determined.

Since we now assume the ground state density to arise from the solution of a non-interacting system, we can write $\rho_0(\mathbf{x}) = \sum_{i=1}^{N} |\phi_i(\mathbf{x})|$. Finding the ground state density then amounts to solving the set of coupled equations for the one-electron system:

$$\left[-\frac{1}{2}\nabla^2 + v^{KS}[\rho](\mathbf{x})\right]\phi_i(\mathbf{x}) = \epsilon_i\phi_i(\mathbf{x}) \quad \text{and} \quad \rho(\mathbf{x}) = \sum_{i=1}^{N}|\phi_i(\mathbf{x})| \tag{3.4.1}$$

But how do we approximate $v^{KS}[\rho]$? Consider the energy functionals for the non-interacting system and the interacting system respectively. If we introduce the *Kohn-Sham kinetic energy*,

$$\mathsf{T}^{KS}[\rho] = -\frac{1}{2}\sum_{i=1}^{N}\int d^3x\ \phi_i^*(\mathbf{x})\nabla^2\phi_i(\mathbf{x}) \tag{3.4.2}$$

then we can write

$$\begin{aligned} E^{KS}[\rho] &= \mathsf{T}^{KS}[\rho] + \int d^3x\ \rho(\mathbf{x})v^{KS}[\rho](\mathbf{x}) \\ E[\rho] &= F[\rho] + \int d^3x\ \rho(\mathbf{x})v(\mathbf{x}) \end{aligned} \tag{3.4.3}$$

We now introduce the *exchange-correlation* energy, defined as

$$E^{XC}[\rho] := F[\rho] - \frac{1}{2}\iint d^3x\ d^3x'\ \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} - \mathsf{T}^{KS}[\rho] \tag{3.4.4}$$

This is a much smaller quantity than $F[\rho]$, since we have subtracted the majority of the kinetic energy as well as the "averaged" electron repulsion energy due to the density. We now want the effective potential $V^{KS}(\mathbf{x})$ corresponding to $\mathsf{W} + V(\mathbf{x})$. Setting $E^{KS}[\rho] = E[\rho]$ in Equation (3.4.3) yields:

$$\begin{aligned} \int d^3x\ \rho(\mathbf{x})v^{KS}[\rho](\mathbf{x}) &= F[\rho] + \int d^3x\ \rho(\mathbf{x})v(\mathbf{x}) - \mathsf{T}^{KS}[\rho] \\ &= \int d^3x\ \rho(\mathbf{x})v(\mathbf{x}) + \frac{1}{2}\iint d^3x\ d^3x'\ \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} + E^{XC}[\rho] \end{aligned} \tag{3.4.5}$$

The potential is then obtained by taking the functional derivative with respect to $\rho(\mathbf{x})$ on both sides of Equation (3.4.5):

$$\begin{aligned} v^{KS}[\rho](\mathbf{x}) &= \frac{\delta}{\delta\rho(\mathbf{x})} \int d^3x\, \rho(\mathbf{x}) v^{KS}[\rho](\mathbf{x}) \\ &= \frac{\delta}{\delta\rho(\mathbf{x})} \left(\int d^3x\, \rho(\mathbf{x}) v(\mathbf{x}) + \frac{1}{2} \iint d^3x\, d^3x'\, \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} + E^{XC}[\rho] \right) \\ &= v(\mathbf{x}) + \frac{1}{2} \int d^3x'\, \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} + v^{XC}[\rho](\mathbf{x}) \end{aligned} \tag{3.4.6}$$

We are thus left with only one quantity to approximate: The exchange-correlation potential $V^{XC}[\rho](\mathbf{x})$, which we expect to be small. Everything else is exact. The exchange correlation energy is comprised of two terms, the residual kinetic energy T^{XC} and the residual electron interaction W^{XC}:

$$\begin{aligned} E^{XC}[\rho] &= \mathsf{T}^{XC}[\rho] + \mathsf{W}^{XC}[\rho] \\ \mathsf{T}^{XC}[\rho] &= \mathsf{T}[\rho] - \mathsf{T}^{KS}[\rho] \\ \mathsf{W}^{XC}[\rho] &= \min_{\Psi\to\rho} \langle \Psi | \mathsf{W} | \Psi \rangle - \frac{1}{2} \iint d^3x\, d^3x'\, \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} \end{aligned} \tag{3.4.7}$$

We can now, finally, find the ground state density of the *interacting* system by solving the following set of coupled equations, called the *Kohn-Sham equations*:

$$\begin{aligned} \left[-\frac{1}{2}\nabla^2 + v^{\text{ext}}(\mathbf{x}) + \int d^3x'\, \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} + v^{XC}[\rho](\mathbf{x}) \right] \phi_i(\mathbf{x}) &= \epsilon_i \phi_i(\mathbf{x}) \\ \sum_{i=1}^{N} |\phi_i(\mathbf{x})|^2 &= \rho(\mathbf{x}) \end{aligned} \tag{3.4.8}$$

The set of coupled equations are usually solved iteratively using the *self consistent field* (SCF) approximation. We will not touch on theoretical convergence properties here, but only note that the systems that have been studied in the present work generally either converge within a roughly constant number of iterations, or do not converge at all.

The Kohn-Sham equations (3.4.8) are accurate to the extent that the potential $V^{XC}[\rho](\mathbf{x})$ is accurate. However, the exact form of the potential is not known, nor is it even known to be computable. Nonetheless, because it is small, we can obtain decent results even with very crude approximations.

Computing good approximations to the true exchange-correlation potential is an entire field unto itself and lies outside the scope of this thesis. Since the introduction of the method in 1965 by Kohn and Sham, a large hierarchy of approximate exchange-correlation potentials have been developed, allowing for tradeoffs between simplicity and efficiency on the one hand, and physicality and accuracy on the other. In the implementation of the current work, we have used the very simplest of these: the non-relativistic, no-spin *local density approximation* (LDA). The method developed in this thesis, though, is agnostic towards the particular exchange-correlation potential that is used, and the XC-potential can be changed in the software by changing a function call.

CHAPTER 4

A linear scaling finite-difference DFT

The finite element DFT method described in this part of the thesis is a more or less straight forward extension of the SIESTA method, which is described in Soler et al. [2002]. This chapter describes the SIESTA method as presented by Soler et al. In essense, we simply summarize the relevant parts of Soler et al. [2002] in preparation for Chapter 5. While everything in the current chapter is covered in Soler et al. [2002], the material is necessary for context.

4.1 Pseudopotentials

The $\mathcal{O}(N)$ scaling property of the SIESTA method relies on the use of numerical basis functions with compact support optimized for pseudopotentials that yield very short-range orbitals and sparse systems of equations. We use norm-conserving pseudopotentials that eliminate the cusps at nuclei as well as oscillatory behaviour. This reduces the need for the extremely fine mesh representations near the nuclei that is otherwise needed to reproduce cusps.

In the nonlocal pseudopotential approximation, the pseudopotential v^{PS} plays the role of the external potential[1] v^{ext} in the Kohn-Sham Equation (3.4.8). That is, the Hamiltonian is

$$\mathsf{H} = -\tfrac{1}{2}\Delta + \mathsf{V}^{PS} + V^H(\mathbf{x}) + V^{XC}[\rho](\mathbf{x}) \tag{4.1.1}$$

where $v^H(\mathbf{x}) = \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}$ is the Hartree potential.

Before we move on, we wish to direct the readers attention to the distinction between the N-electron non-interacting Hamiltonian H, which must be used when obtaining the total energy, and the one-electron Hamiltonian h from which it is built, and which is the operator that is diagonalized in the Kohn-Sham Equation (3.4.8). That is, we have

$$\mathsf{H} = \bigwedge^{N} \mathsf{h} \tag{4.1.2}$$

and

$$\mathsf{h} = -\tfrac{1}{2}\nabla^2 + \mathsf{v}^{PS} + v^H(\mathbf{x}) + v^{XC}[\rho](\mathbf{x}) \tag{4.1.3}$$

Mixing up the two Kohn-Sham Hamiltonians will inevitably lead to a general state of confusion.

[1] Recalling the convention introduced in Chapter 3 of using upper case letters for N-electron potentials and the corresponding lowercase letters for their one- or two-electron constituent parts. For example, $V^H(\mathbf{x}) = \sum_i v^H(\mathbf{x}_i)$. This necessitates a divergence from the notation in Soler et al. [2002], in which upper- or lowercase is used at random. Other than this small difference, the notation is kept the same.

Like Soler et al., we use Troullier-Martins parameterized pseudopotentials, transformed into a fully non-local form proposed by Kleinman and Bylander (KB) Kleinman and Bylander [1982]:

$$\begin{aligned} \mathsf{v}^{PS} &= v_{local}(\mathbf{x}) + \mathsf{v}^{KB} \\ \mathsf{v}^{KB} &= \sum_{l=0}^{l_{\max}^{KB}} \sum_{m=-l}^{l} \sum_{n=l+1}^{N_l^{KB}} \left|\chi_{nlm}^{KB}\right\rangle v_{nl}^{KB} \left\langle\chi_{nlm}^{KB}\right| \\ v_{ln}^{KB} &= \langle\varphi_{nl}|\delta v_l|\varphi_{nl}\rangle \\ \delta v_l(r) &= v_l(r) - v_{local}(r) \end{aligned} \tag{4.1.4}$$

The functions

$$\chi_{nlm}^{KB}(\mathbf{x}) = \delta v_l(r)\varphi_{nl}(r)Y_{lm}(\hat{\mathbf{x}}) \tag{4.1.5}$$

are called the *KB projection functions*. We will index nuclei with upper case I and J.

In order to eliminate the long range of $v_I^{local}(\mathbf{x})$, we screen it with the potential v_I^{atom} generated by an appropriate approximate atomic electron density ρ_I^{atom}. The resulting screened *neutral atom potential* $v^{NA} = v_I^{local} + v_I^{atom}$ is zero outside of the cutoff radius r_c for the atomic orbitals. We now write

$$\begin{aligned} \delta\rho(\mathbf{x}) &= \rho(\mathbf{x}) - \sum_I \rho_I^{atom}(\mathbf{x}) \\ \delta v^H(\mathbf{x}) &= \frac{\delta\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|} \end{aligned} \tag{4.1.6}$$

whereby the Hamiltonian becomes

$$\begin{aligned} \mathsf{H} &= -\tfrac{1}{2}\Delta + \sum_I (V_I^{local}(\mathbf{x}) + \mathsf{V}_I^{KB}) + V^H(\mathbf{x}) + V^{xc}(\mathbf{x}) \\ &= \tfrac{1}{2}\Delta + \sum_I \mathsf{V}_I^{KB} + \sum_I V_I^{NA}(\mathbf{x}) + \delta V^H(\mathbf{x}) + V^{XC}[\rho](\mathbf{x}) \end{aligned} \tag{4.1.7}$$

4.2 Numerically optimized orbitals with finite support

The atomic orbitals are one-electron basis functions of the form

$$\begin{aligned} \phi_{Ilmn}(\mathbf{x}) &= \phi_{lmn}(\mathbf{x} - \mathbf{X}_I) \\ \phi_{lmn}(\mathbf{x}) &= \phi_{ln}(r)Y_{lm}(\hat{\mathbf{x}}) \end{aligned} \tag{4.2.1}$$

Note that the index n here doesn't denote the number of nodes in the radial function, contrary to how it will be used in the later two parts of this thesis. It is merely an index labelling several orbitals that have the same angular dependence Y_{lm}, but different radial parts. The radial functions are, very roughly speaking, solutions to approximate one-electron Schrödinger equations of the atomic pseudopotentials $v_l(r)$, which ensure smooth convergence to zero at the cutoff radii r_c. The exact means by which the numerical radial functions are determined from the atomic pseudopotentials $v_l(r)$ are described in Section III of Soler et al. [2002]. Examples of the radial parts $\phi_{ln}(r)$ of computed numerical orbitals are given in Figure 4.1.

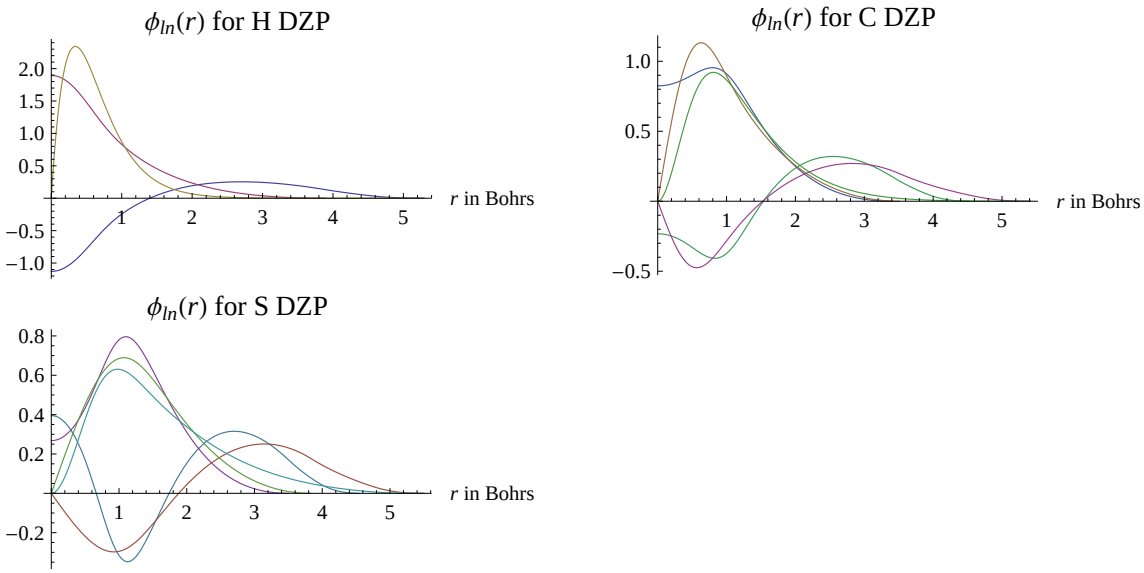


Figure 4.1: Radial parts $\phi_{ln}(r)$ of double-ζ polarized (DZP) numerical orbitals for hydrogen, carbon and sulfur.

4.3 Computational parts comprising the SCF loop

4.3.1 Constructing the electron Hamiltonian

Two-center integrals

We evaluate the two-center overlap integrals by using the Convolution Theorem and isometry of the Fourier transform, obtaining

$$\mathsf{S}(\mathbf{X}) \equiv \int d^3x\, \phi_1^*(\mathbf{x})\phi_2(\mathbf{x}-\mathbf{X}) = \int d^3p \phi_1^{t*}(\mathbf{p})\phi_2^t(-\mathbf{p})e^{-i\mathbf{p}\cdot\mathbf{X}} \tag{4.3.1}$$

If, as is the case for the orbitals ϕ_{nlm} and the KB projections χ^{KB}_{nlm}, the ϕ's only contain a single spherical harmonic, their Fourier transforms are of the form

$$\phi^t(\mathbf{p}) = \tilde{\phi}(p)Y_{lm}(\hat{\mathbf{p}}) \tag{4.3.2}$$

and we can expand the plane wave in terms of spherical harmonics and spherical Bessel functions

$$e^{-i\mathbf{p}\cdot\mathbf{X}} = 4\pi \sum_{l=0}^{\infty}\sum_{m=-l}^{l} (-i)^l j_l(pR)Y_{lm}(\hat{\mathbf{p}})Y^*_{lm}(\hat{\mathbf{X}}) \tag{4.3.3}$$

to obtain

$$\begin{aligned}
\int d^3p\, \phi_1^{t*}(\mathbf{p})\phi_2^t(\mathbf{p})e^{-i\mathbf{p}\cdot\mathbf{X}} &= 4\pi \sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y^*_{lm}(\hat{\mathbf{X}}) \int d^3p\, (-i)^l j_l(pR)\phi_1^{t*}(\mathbf{p})\phi_2^t(\mathbf{p})Y_{lm}(\hat{\mathbf{p}}) \\
&= 4\pi \sum_{l=0}^{\infty}\sum_{m=-l}^{l} (-i)^l \int_0^{\infty} dp\, p^2 j_l(pR)\tilde{\phi}_1^*(p)\tilde{\phi}_2(p) \\
&\times \int d\Omega\, Y^*_{l_1m_1}(\hat{\mathbf{p}})Y_{l_2m_2}(\hat{\mathbf{p}})Y_{lm}(\hat{\mathbf{p}}) \\
&= 4\pi \sum_{l=|l_1-l_2|}^{l_1+l_2}\sum_{m=-l}^{l} c^{lm}_{l_1m_1,l_2,m_2} S_{lm}(R)Y_{lm}(\hat{\mathbf{x}})
\end{aligned} \tag{4.3.4}$$

where

$$\begin{aligned} S_l(R) &\equiv (-i)^l \int_0^\infty dp\ p^2 \tilde{\phi}_1^*(p)\tilde{\phi}_2(p) j_l(pR) \\ c^{lm}_{l_1 m_1, l_2, m_2} &\equiv \int d\Omega Y^*_{l_1 m_1}(\hat{\mathbf{p}}) Y_{l_2 m_2}(\hat{\mathbf{p}}) Y_{lm}(\hat{\mathbf{p}}) \end{aligned} \tag{4.3.5}$$

The angular integrals $c^{lm}_{l_1 m_1, l_2, m_2}$ can easily be derived using either the harmonic projection methods described in Section 10.3 or using Clebsch-Gordan coefficients, or even obtained numerically, as is done in Soler et al. [2002]. The radial integrals $S_l(R)$ involve numerical functions. However, since they are one-dimensional and cut off at $r_1^c + r_2^c$, they can be obtained accurately and quickly by numerical integration. Both the angular and radial integrals can be computed once and for all for a given basis and stored. The kinetic matrix elements

$$\mathsf{T}(\mathbf{X}) \equiv -\frac{1}{2}\int d^3x\ \phi_1^*(\mathbf{x})\nabla^2\phi_2(\mathbf{x}-\mathbf{X}) = \frac{1}{2}\int d^3p\ p^2 \phi_1^{*t}(\mathbf{p})\phi_2^t(\mathbf{p}) e^{-i\mathbf{p}\cdot\mathbf{X}} \tag{4.3.6}$$

are computed in an analogous fashion, only with $S_l(R)$ replaced by $T_l(R)$, which has an extra factor k^2:

$$T_l(R) \equiv i^l \int_0^\infty dp\ p^4 \tilde{\phi}_1^*(p)\tilde{\phi}_2(p) j_l(pR) \tag{4.3.7}$$

Real space integrals

The final three terms in Equation (4.1.7) involve potentials that are calculated point-wise. The short-range neutral-atom pseudopotentials $v_I^{NA}(r)$ are spherically symmetric and can be tabulated on a fine radial grid and interpolated. The remaining two terms require calculating the electron density $\rho(\mathbf{x})$ and difference density $\delta\rho(\mathbf{x})$ on the grid. In addition, in order to construct the Hartree difference potential

$$\delta v^H(\mathbf{x}) = \frac{\delta\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}$$

we must solve Poisson's Equation for the difference density $\delta\rho(\mathbf{x})$:

$$-4\pi\epsilon(\mathbf{x})\nabla^2(\delta v^H(\mathbf{x}) + v^{\text{ext}}(\mathbf{x})) = \delta\rho(\mathbf{x}) \tag{4.3.8}$$

4.3.2 Constructing the density matrix and the density

Let ψ_i be the Kohn-Sham eigenstates, i.e. the eigenfunctions of the Kohn-Sham Hamiltonian, and

$$\psi_i(\mathbf{x}) = \sum_\tau \phi_\tau(\mathbf{x}) c_{i,\tau} \tag{4.3.9}$$

where $\tau = (I, l, m, n)$ includes both the atomic- and the orbital index. The ground state density is then[2]

$$\rho(\mathbf{x}) = \sum_{i=1}^N |\psi_i(\mathbf{x})|^2 \tag{4.3.10}$$

Then the density matrix $\mathbf{D}$ is defined by

$$D_{\tau'\tau} = \sum_{i=1}^N c^*_{i,\tau'} c_{i,\tau} = \sum_{i=1}^N \langle\phi_{\tau'}|\psi_i\rangle\langle\psi_i|\phi_\tau\rangle \tag{4.3.11}$$

[2]If spin is not included in the Hamiltonian, then each ψ_i must appear twice.

and the ground state density is

$$\rho(\mathbf{x}) = \sum_{\tau',\tau} \phi^*_{\tau'}(\mathbf{x}) D_{\tau'\tau} \phi_\tau(\mathbf{x}) \tag{4.3.12}$$

Although this operation, as written, looks like a $\mathcal{O}(N^2_{\text{orbs}})$ operation in the number N_{orbs} of atomic orbitals, sparsity makes it an $\mathcal{O}(N_{\text{orbs}})$ operation.

4.4 Computing the total energy

The Kohn-Sham total energy can be written as the *band-structure* energy minus a sum of *double count* corrections. The band-structure energy is the trace of the Hamiltonian times the density matrix:

$$E^{BS} = \sum_{i=1}^{N} n_i \langle \psi_i | \mathsf{H} | \psi_i \rangle = \sum_{i,j} D_{ij} H_{ji} = Tr(\mathsf{HD}) \tag{4.4.1}$$

Adding the correction terms, we get the following expression for the total Kohn-Sham energy:

$$E^{KS} = Tr(\mathsf{HD}) - \frac{1}{2}\int d^3x\, v^H(\mathbf{x})\rho(\mathbf{x}) + \int d^3x\, \rho(\mathbf{x})\, (\epsilon^{xc}(\mathbf{x}) - v^{XC}[\rho](\mathbf{x})) + \sum_{I<J} \frac{Z_I Z_J}{R_{IJ}} \tag{4.4.2}$$

When we expand the terms in Equation (4.4.2), we obtain (Soler et al. [2002], Equation (57)):

$$\begin{aligned} E^{KS} = Tr((\mathsf{T} + \mathsf{V}^{KB})\mathsf{D}) + \sum_{I<J} U^{NA}_{IJ}(R_{IJ}) + \sum_{I<J} \delta U^{local}_{IJ}(R_{IJ}) + \sum_I U^{atom}_I \\ + \int d^3x\, v^{NA}(\mathbf{x})\delta\rho(\mathbf{x}) + \frac{1}{2}\int d^3x\, \delta v^H(\mathbf{x})\delta\rho(\mathbf{x}) + \int d^3x\, \epsilon^{xc}(\mathbf{x})\rho(\mathbf{x}) \end{aligned} \tag{4.4.3}$$

where T is the kinetic energy operator. The atomic interaction terms are described in Soler et al. [2002], pages 2759–2760. However, for the purpose of this thesis, we are only interested in the final three terms, since these are the only ones affected when lifting the real space representation from grids to finite element spaces.

In all, the following are the individual terms in the total energy that we must compute using finite element representations:

$$\begin{aligned} E^{\delta\text{NA}} &= \int d^3x\, v^{NA}(\mathbf{x})\delta\rho(\mathbf{x}) \\ E^{\delta\text{H}} &= \frac{1}{2}\int d^3x\, \delta v^H(\mathbf{x})\delta\rho(\mathbf{x}) \\ E^{\text{xc}} &= \int d^3x\, \epsilon^{xc}[\rho](\mathbf{x})\, \rho(\mathbf{x}) \\ E^{\text{ext}} &= \int d^3x\, v^{\text{ext}}(\mathbf{x})\rho(\mathbf{x}) - \sum_I v^{\text{ext}}(\mathbf{R}_I) Z_I \end{aligned} \tag{4.4.4}$$

In the next chapter, we will describe the changes necessary to lift the method to a finite element setting.

Chapter 5

Adapting the DFT-method to the finite-element method

5.1 Overview

Figure 5.1 shows the computational parts of the Kohn-Sham self-consistent field loop. The steps that must be changed to accommodate the finite element method are outlined. These steps are detailed in Section 5.4. We must make sure that every operation in the SCF-loop remains at most $\mathcal{O}(N_{\text{orbs}})$ operations (where N_{orbs} in the total number of orbitals) under the same restrictions that

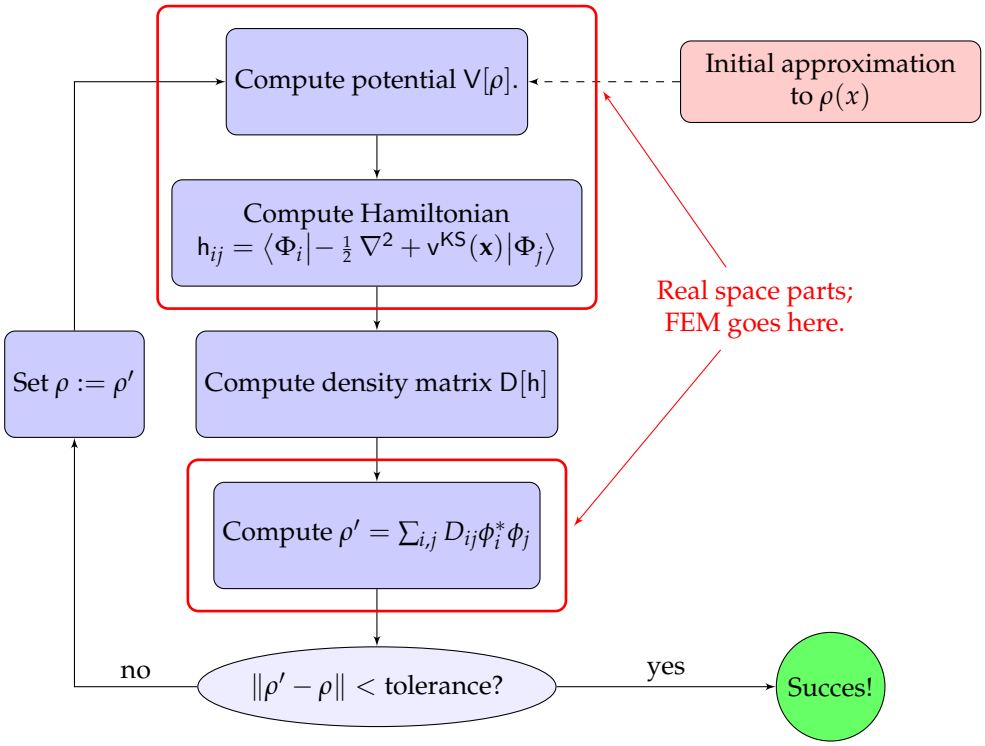


Figure 5.1: The DFT self consistent loop

ensure linearity of the SIESTA method, i.e. when we bound by a constant the average number of orbitals overlapping at each point in space.

5.2 Dual function representations

The finite element method is essentially a Sobolev-space method: While functions in finite element spaces do have well-defined point-wise values, evaluating a finite element function at an arbitrary point x is not a fast operation: First, the cell in which the point resides must be found, then the corresponding unit cell coordinate $\xi = \chi_c^{-1}(x)$, and finally the local nodes e_i must be evaluated at ξ:

$$f(x) = \sum_{i=1}^{n_{\text{dof}}} a_i e_i(\xi)$$

In DFT we not only need the electron density as a distribution, but also require point-wise values in order to produce the exchange-correlation integrals. In addition, for the present DFT method, we must add core-charge and compensation-charge (see Section 4.3.1) point-wise.

To solve this problem, we work with two representations of functions:

1. Finite element functions (*FE-functions*), represented as the coefficients $(a_1, \ldots, a_{N_{\text{dof}}})$ in the expansion

   $$f = \sum_{i=1}^{N_{\text{dof}}} a_i \mathfrak{e}_i$$

   This form is used in the solution of the Poisson equation, and for all operations that do not rely on point-wise evaluation.

2. Point-wise defined functions (*Point-functions*), represented by the values $(f(x_1), \ldots, f(x_{N_q}))$ at every quadrature point. We can equivalently write this as an $N_{\text{cells}} \times n_q$ matrix

   $$\mathbf{F}_{cq} = f(x_{cq})$$

   Note that these are the same numbers in the same order, only indexed differently: $f(x_{cq}) = f(x_{(c-1)n_q+q})$.

The restriction to evaluation at quadrature points is key, as we will see in the following subsection. In addition, if a sufficiently fine quadrature formula is chosen, FE-functions and point-functions are equivalent. That is, n_q must be large enough that integrals are computed exactly for products of two finite element functions. If this is not the case, some information is lost when converting between the representations.

The astute reader may wonder why we do not instead define point-functions at nodal positions. In this case, evaluation is trivial:

$$f = \sum_{i=1}^{N_{\text{dof}}} a_i \mathfrak{e}_i \iff \mathbf{F}_i = a_i$$

that is, the map between FE-functions and point-functions is the identity, since the point-wise values are the same as the coefficients. However, while using the nodal positions eliminates the need for a dual representation, we must use the quadrature points (with their appropriate weights) when integrating: the nodal positions offer no guarantee of accuracy when integrating by quadrature, whereas the quadrature points ensure that functions up to a certain polynomial degree are integrated exactly.

5.2.1 Converting between FE-functions and point-wise function values

Let us assume that we have a point-wise representation $\mathbf{F}$ of the function f. We now wish to find a finite element function $\tilde{f} \in \tilde{V}$ that obeys

$$\langle \tilde{f} | v \rangle = \langle f | v \rangle \quad \forall v \in \tilde{V} \tag{5.2.1}$$

If our quadrature formula is sufficiently fine, we have for the integral on a single cell c:

$$\int_c dx\, f(x) e_i(\mathsf{x}_c^{-1}(x)) = \sum_{q=1}^{n_q} f(x_{cq}) e_i(\xi_q) w_q \tag{5.2.2}$$

Otherwise, the equality is only approximate. We can rewrite Equation (5.2.2) to

$$\int_c dx\, f(x) e_i(\mathsf{x}_c^{-1}(x)) = \mathbf{Q}_i \cdot \mathbf{F}_c, \quad \mathbf{Q}_{iq} := e_i(\xi_q) w_q \tag{5.2.3}$$

where $F_{cq} = f(\mathsf{x}_c(\xi_q))$. If we now consider the restriction of $\tilde{f}$ to c

$$\tilde{f}_c = \sum_{i=1}^{n_{\text{dof}}} a_{ci} e_i \tag{5.2.4}$$

we obtain

$$\mathbf{M}\mathbf{a}_c = \mathbf{Q}\mathbf{F}_c \implies \mathbf{a}_c = \mathbf{M}^{-1}\mathbf{Q}\mathbf{F}_c \tag{5.2.5}$$

The *unit cell projection matrix* $\mathbf{X} = \mathbf{M}^{-1}\mathbf{Q}$ is a small $n_{\text{dof}} \times n_q$ matrix. It is the same for all the cells in a mesh, and it does not change.

Now, it is easy to see that we can perform the converse operation, i.e. interpolating the finite-element function onto the quadrature points, by applying the constant $n_q \times n_{\text{dof}}$ matrix $\mathbf{X}^{-1} = \mathbf{Q}^{-1}\mathbf{M}$:

$$\mathbf{Q}^{-1}\mathbf{M}\mathbf{a}_c = \mathbf{F}_c \tag{5.2.6}$$

5.3 Distributed real space representation of the orbitals

in Section 4.3, we needed two representations of atomic orbitals: As n_{orbs} radial functions times spherical harmonic used for two-center integrals, and as N_{orbs} representations on the real-space grid.

Everything involving only the first representation remains the same when adding the finite element method. For the real-space representation, it has proved most convenient to represent the one-electron basis functions as a sparse $N_q \times N_{\text{orbs}}$ matrix $\Phi_{qi} = \phi_i(x_q)$. While Φ is sparse, the number of nonzero elements per row bounded by the number of orbitals defined at a given point, the size of Φ can be quite large for fine systems. We therefore distribute the matrix, dividing calculation and storage of the matrix among computational nodes, making each node responsible only for a subset of the quadrature points.

5.4 Computational steps in the SCF loop

5.4.1 Computing the density

The density matrix $\mathbf{D}$ is computed the same way as before, since it does not involve real-space function representations.

Listing 5.1: Computing a `FEFunction` from a `PointFunction`.

```
ConstructFEFunction(const PointFunction& fp, FEFunction& f) const
{
  FETools::compute_projection_from_quadrature_points_matrix
    (fe,quadrature_formula,quadrature_formula,X);
  {
    typename DoFHandler<dim>::active_cell_iterator
      cell = dof_handler.begin_active(),
      endc = dof_handler.end();

    Vector<double> U(n_q_pts), V(n_cell_dof);
    std::vector<unsigned int> global_dof_indices (n_cell_dof);

    for (size_t pt=0; cell!=endc; ++cell, pt += n_q_pts){
      cell->get_dof_indices (global_dof_indices);

      for(size_t i=0;i<n_q_pts;i++) U[i] = fp[pt+i];

      projection_matrix_X.vmult(V,U);
      for(size_t i=0;i<n_cell_dof;i++)
        f[global_dof_indices[i]] = V[i];
    }
  }
}
```

The density as a function of $\mathbf{x}$ is obtained from the density matrix as follows:

$$\rho(\mathbf{x}) = \sum_{i,j=1}^{N_{\text{orbs}}} \phi_i(\mathbf{x}) D_{ij} \phi_j(\mathbf{x}) \tag{5.4.1}$$

On its face, constructing ρ looks like an $\mathcal{O}(N_q\ N_{\text{orbs}}^2)$ operation. However, we can expose the sparsity by writing it in the equivalent form:

$$\rho(x_q) = \sum_{i,j \in idx(\Phi_q)} \Phi_{qi} D_{ij} \Phi_{qj} \tag{5.4.2}$$

The row Φ_q is sparse, with the length $|idx(\Phi_q)| = n_{\text{orbs}}(x_q)$ the number of orbitals with support at the point x_q. If

$$\underset{x_q}{\text{average}}\ n_{\text{orbs}}(x_q)^2 = \frac{1}{N_q} \sum_{q=1}^{N_q} |idx(\Phi_q)|^2 \tag{5.4.3}$$

is bounded from above by a constant as we grow the system, then computing ρ is a time $\mathcal{O}(N_q)$ operation, i.e. asymptotically dependent only on the number of spatial points to be evaluated.

5.4.2 Computing matrix representations

The inner product, weighted by the potential $v(\mathbf{x})$, is given by

$$\langle \phi_i | v(\mathbf{x}) | \phi_j \rangle = \int_\Omega d^3x\ \phi_a(\mathbf{x}) v(\mathbf{x}) \phi_b(\mathbf{x}) \approx \sum_{q=1}^{N_q} w_{q \bmod n_q}\ \Phi_{qi} v_q \Phi_{qj} \tag{5.4.4}$$

where in the right hand side, we use the point-function representation $v_q = v(x_q)$. However, Equation (5.4.4) is quite inefficient due to the distributed sparse structure of Φ_{qi}.

Listing 5.2: Computing the density `PointFunction` using distributed sparse matrices.

```
void PFBasis::CalculateDensity(Epetra_CrsMatrix& DensityMatrix, PointFunction& density) const {
  const size_t num_max_i = DensityMatrix.MaxNumEntries();
  const size_t num_my_x = Phi->NumMyRows();

  // This processor owns only a limited region of space
  for(size_t xLocal=0;xLocal<num_my_x;xLocal++){
    int x = Phi->GRID(xLocal); // Global index of xLocal
    double density_x = 0;

    int indices_phi[num_max_i], num_i;
    double phi_x [num_max_i];

    // Indices: [i_1,...,i_num_i], values: phi_xj = phi_ij(x)
    Phi->ExtractGlobalRowCopy(x,num_max_i,num_i,phi_x,indices_phi);

    for(size_t i=0;i<num_i;i++){
      double Di[num_max_i];
      int indices_Di[num_max_i], num_Di;

      // Extract all nonzero D_ij in row D_i:
      DensityMatrix.ExtractGlobalRowCopy(indices_phi[i],num_max_i,num_Di,Di,indices_Di);

      // rho(x) = sum_ij = sum_ij phi_xi D_ij phi_xj = sum_i phi_xi (phi_x . D_i)
      density_x += phi_x[i]*dotProduct(phi_x,indices_phi,num_i, Di, indices_Di, num_Di);
    }
    // Write to global structure; No contention, since we own x
    density[x] = density_x;
  }
}
```

This following reorganization of the sum solves this problem. Let $\mathbf{V}$ be the resulting matrix representation for v. For each point x_q:

1. Get indices $idx(\Phi_q)$ of orbitals with non-zero value at x_q.
2. For each i:
   a) Compute the sparse row $r_j = \Phi_{qi} v(x_q) \Phi_{qj}$ for $j \in idx(\Phi_q)$.
   b) Add the row $\mathbf{r}$ to $\mathbf{V}_i$.

5.5 Computing the Hamiltonian

We first compute the external potential $v^{\text{ext}}(\mathbf{x})$ and the difference Hartree potential $\delta v^H(\mathbf{x})$ The external potential is computed by solving Poisson's equation in the environment without any electron density.

$$\begin{aligned} -4\pi\epsilon(\mathbf{x})\nabla^2 v^{\text{ext}}(\mathbf{x}) &= 0 \\ -4\pi\epsilon(\mathbf{x})\nabla^2(\delta v^H(\mathbf{x}) + v^{\text{ext}}(\mathbf{x})) &= \delta\rho(x) \end{aligned} \tag{5.5.1}$$

We can then separate the external potential and the difference Hartree potential by subtracting $v^{\text{ext}}(\mathbf{x})$ from the solution to the second Poisson equation in (5.5.1). The remaining real-space terms

Listing 5.3: Computing the Hamiltonian from the density matrix.

```
void LatticeCalculator::CalculateHamiltonian(SparseMatrix& h)
{
    //kinetic term
    h=tBSystem->getH0();
    tBSystem->AddHnl(QFEMORB_NS::GetNonLocalAtomTerm(basisSet),h);

    //******** calculate the effective potential ******//
    //Full real space density ρ(x) needed to compute v^XC
    PointFunction rho_pt, deltarho_pt;
    BasisOrbitals.CalculateDensity(densityMatrix,rho_pt);

    PointFunction vEff_pt;
    FEFunction vEff_fe;
    {
      // Screened density δρ(x)
      PointFunction deltarho_pt(rho_pt);
      deltarho_pt += CompensationCharge(basisSet,space);

      // Solve Poissons equation using libspace to get δv^H(x) + v^ext(x)
      space.SolvePoisson(deltarho_pt*(4*M_PI), vEff_fe);
      space.ConstructPointFunction(vEff_fe,vEff_pt);

      rho_pt += CoreCharge(basisSet,space);
      vEff_pt += ExchangeCorrelation(basisSet.XCFunctionalType(), space, rho_pt,true);
    }

    // Add local potential v^NA(x)
    vEff_pt += VZero(basisSet,space);

    Epetra_CrsMatrix hDum(h);
    BasisOrbitals.EvaluateMatrix(vEff_pt,hDum);
    QMATH_NS::Add(hDum,false,1.0, h,1.0);
}
```

are computed and added point-wise to the point-function representation of $\delta v^H(\mathbf{x})$:

$$v^{eff}(\mathbf{x}) = v^{XC}[\rho](\mathbf{x}) + \delta v^H(\mathbf{x}) + v^{ext}(\mathbf{x}) + v^{NA}(\mathbf{x}) \tag{5.5.2}$$

whereby the one-electron Kohn-Sham Hamiltonian can be computed:

$$\mathsf{h}_{ij} = \mathsf{h}^0_{ij} + \left\langle \phi_i \middle| \mathsf{v}^{eff} \right\rangle \phi_j \tag{5.5.3}$$

Listing 5.3 shows how the part of the Hamiltonian that depends on the finite element representation is calculated.

5.6 Computing total energy

We recall that the terms in the total energy computation that require special treatment in the finite-element setting are the following:

$$\begin{aligned} E^{\delta\mathrm{NA}} &= \int d^3x\, v^{NA}(\mathbf{x})\delta\rho(\mathbf{x}) \\ E^{\delta\mathrm{H}} &= \frac{1}{2}\int d^3x\, \delta v^{H}(\mathbf{x})\delta\rho(\mathbf{x}) \\ E^{\mathrm{xc}} &= \int d^3x\, \epsilon^{xc}[\rho](\mathbf{x})\, \rho(\mathbf{x}) \\ E^{\mathrm{ext}} &= \int d^3x\, v^{\mathrm{ext}}(\mathbf{x})\rho(\mathbf{x}) - \sum_I v^{\mathrm{ext}}(\mathbf{R}_I)Z_I \end{aligned} \tag{5.6.1}$$

The rest are calculated the same as before. Listing 5.4 shows in detail, how the individual energy contributions $E^{\delta H}$, E^{XC}, $E^{\delta NA}$, and E^{ext} are computed.

5.7 Refining the Mesh: Constructing good finite-element spaces

Assuming the preconditions in Theorem 2.3.3 are met, the finite element solution $v_h(\mathbf{x})$ to Poisson's equation obeys

$$\|v - v_h\|_{L^2} \leq C \left\| \sqrt{\epsilon} h^2 |\nabla^2 v| \right\|_{L^2} \tag{5.7.1}$$

(where for notational convenience, we write $v(\mathbf{x}) = \delta v^H + v^{\mathrm{ext}}(\mathbf{x})$) Thus, to decrease the error of the computed Hartree potential, we need to subdivide cells for which the right hand side is large.

However, the goal of computing faithful approximations v_h to $v(\mathbf{x})$ is only a sub-goal of our greater scheme: To compute approximations to the ground state density $\rho(\mathbf{x})$ and approximations to the density functionals that describe the system's observables.

The functional that determines the ground state solution $\rho(\mathbf{x})$ is the total energy, and the total energy and its constituent terms also form the observables in which we are the most interested. All the energy integrals are weighted point-wise by either the electron density $\rho(\mathbf{x})$ or the difference density $\delta\rho(\mathbf{x})$. The energies involving the Poisson solution give rise to the errors

$$\begin{aligned} \mathrm{error}(E^{\delta H}) &= \int_\Omega d^3x\, \delta\rho(\mathbf{x})\, (\delta v(\mathbf{x}) - \delta v_h(\mathbf{x})) \\ \mathrm{error}(E^{\delta NA}) &= \int_\Omega d^3x\, \delta\rho(\mathbf{x})\, (v^{NA}(\mathbf{x}) - v_h^{NA}(\mathbf{x})) \\ \mathrm{error}(E^{\mathrm{ext}}) &= \int_\Omega d^3x\, \rho(\mathbf{x})\, (v^{\mathrm{ext}}(\mathbf{x}) - \delta v_h^{\mathrm{ext}}(\mathbf{x})) - \sum_I Z_I (v^{\mathrm{ext}}(\mathbf{X}_I) - v_h^{\mathrm{ext}}(\mathbf{X}_I) \end{aligned} \tag{5.7.2}$$

The contribution of error from a given cell c is then the actual error weighted by $\rho(\mathbf{x})$ or $\delta\rho(\mathbf{x})$: The approximate error in the solution to the potential $v^{\mathrm{eff}}(\mathbf{x})$ outside the support of ρ does not contribute at all to the error in the energy norm. However, the quality of the mesh outside the support of ρ is still important: It must be regular enough that long range effects do not disrupt the solution inside the support of ρ.

Finally, the density must *also* be represented on the finite element space. Thus, there is also error introduced from integration by quadrature in Equation (5.7.2), due to the finite element representation of $\rho(\mathbf{x})$ and $\delta\rho(\mathbf{x})$. To obtain good pointwise representation of $\rho(\mathbf{x})$ and $\delta\rho(\mathbf{x})$ also when they are small (but non-zero), we subdivide according to Step 2 below.

Listing 5.4: Total energy computation of converged system.

```
double LatticeCalculator::HartreeExchangeEnergy(double& eH, double& eVNA, double& eXC, double& eExt) {
  PointFunction rho_pt, vEff_pt, vExt_pt;
  const Epetra_Comm& comm(tBSystem_->getS().Comm());

  BasisOrbitals.CalculateDensity(densityMatrix,rho_pt);

  PointFunction deltarho_pt(rho_pt);
  FEFunction null_fe(rho_pt.size(),0.0), vEff_fe, vExt_fe, rho_fe, deltarho_fe,XC_fe;

  deltarho_pt += CompensationCharge(basisSet,space); // Construct δρ(x) from ρ(x)
  space.ConstructFEFunction(deltarho_pt,deltarho_fe);

  // Solve Poisson's equation for δρ(x) using libspace; Produces δv^H + v^ext
  space.SolvePoisson(deltarho_pt*(4*M_PI),vEff_fe);

  // Compute and subtract v^ext to obtain δv^H
  space.SolvePoisson(null_fe,vExt_fe);
  vEff_fe -= vExt_fe;
  vEff_fe *= 0.5; //multiply δv^H by 0.5 for energy integrals

  eH = space.Integrate(deltarho_fe,vEff_fe); // E^δH = 1/2 ∫_Ω d^3x δv^H(x)δρ(x)

  // Energy from external potential: E^ext = ∫_Ω d^3x v^ext(x)ρ(x) − Σ_I Z_I v^ext(R_I)
  double eExt_centers = 0;
  {
    vector<coordinate> centers(get_centers(*basisSet_));

    for(size_t i=0;i<centers.size();i++){
      double Z_I = basisSet.NumberElectrons(i); // Pseudoatom-charges
      double vExtofX =space.Value(vExt_fe,centers[i]);
      eExt_centers += Z_I*vExtofX;
    }
  }
  eExt = space.Integrate(rho_fe,vExt_fe) - eExt_centers;

  space.increase_quadrature_order(2);
  space.ConstructPointFunction(vEff_fe,vEff_pt);
  space.ConstructPointFunction(rho_fe,rho_pt);

  // vEff is now 1/2 δv^H + v^0
  vEff_pt += VZero(basisSet,space); // Add local potential

  //The electron density used for v^XC(x) includes the core charges
  rho_pt += CoreCharge(basisSet,space);
  PointFunction XC_pt(ExchangeCorrelation(basisSet, space, rho_pt,false));

  space.ConstructFEFunction(vEff_pt,vEff_fe);
  space.ConstructFEFunction(rho_pt, rho_fe);
  space.ConstructFEFunction(XC_pt, XC_fe);
  space.increase_quadrature_order(-2);

  double eHNA = space.Integrate(deltarho_fe,vEff_fe);
  eVNA = eHNA-eH; // Neutral atom-energy

  eXC = space.Integrate(rho_fe,XC_fe); // Exchange-Correlation energy

  return eXC+eHNA;
}
```

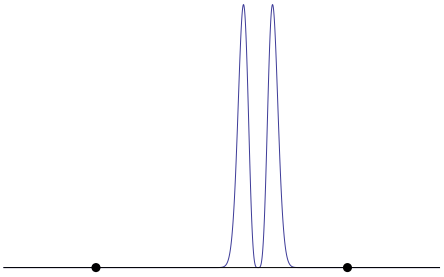


Figure 5.2: A one-dimensional cell with 2-point Gaussian quadrature (corresponding to linear elements). A large amount of electron density is contained in the cell, but the quadrature approximation of the integral is zero, and the cell is not subdivided. .

However, we are faced with a problem: The integrals $\int_c d^3x\, \rho(\mathbf{x})$ must be evaluated by quadrature on an already existing mesh. If a cell is too coarse, we won't catch regions with high density or curvature. Figure 5.2 illustrates this situation.

We can make sure that we catch all cells that contain electron density in the following manner: Orbitals have a fixed range r_{cutoff}. Outside the cutoff range r_{cutoff} of a nucleus I, no atomic orbital centered on I contributes to the electron density. The exact range r_{cutoff} depends on the basis set, but is typically in the order of 3–5Å. We can now successively subdivide cells that are within distance r_{cutoff} of at least one nucleus, until the following relation holds for every cell c and nucleus I:

$$\text{dist}(c, I) > R_{\text{cutoff}} \quad \text{or} \quad \text{diameter}(c) < \max(R_{\text{cutoff}} - \text{dist}(c, I), d_{\min}) \tag{5.7.3}$$

where adding $d_{\min}$ is chosen small enough that electron density will be found, but large enough that we do not obtain overly detailed meshes. We have used $d_{\min} \approx r_{\text{cutoff}}/2$ in this work.

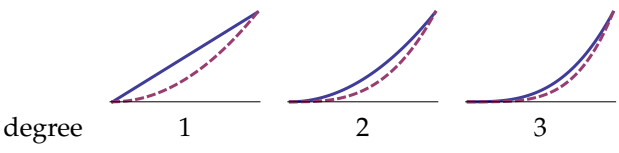
In total, the following scheme is used for adaptive mesh refinement:

1. As an initial step, we recursively subdivide cells that are within max r_{cutoff} of a nucleus with diameter too large according to Equation (5.7.3). This ensures that we can roughly estimate the integrals $\int_c \rho(x) d^3x$ and $\int_c \delta\rho(x) d^3x$ on every cell.

2. We subdivide cells c for which $\int_c \rho(x) d^3x$ or $\int_c \delta\rho(x) d^3x$ is greater than some set amount of electron density. This can be done initially for the zero-order density $\rho_0(\mathbf{x})$, since $\rho_0(\mathbf{x})$ can be calculated without a mesh. The cell's electron content is estimated by quadrature on the cell, and cells are subdivided if
$$\|\rho\|_c > \delta e$$
for a fixed threshold δe.

   This is the most important step, and leads to the greatest detail.

3. We subdivide cells where $\rho(x)$ or $\delta\rho(\mathbf{x})$ has "too much curvature". This reduces the L^2-error of the finite element representation. By "too much curvature" we mean the cell-norm of the derivative of degree one higher than the finite element polynomial degree d. This is a good estimate of the finite element representation's inaccuracy: Any function that is up to a degree-d polynomial on the cell c is represented exactly on c. The L^2-error stems from the higher terms, as illustrated below:



The curvature is estimated numerically, and cells with degree k elements are subdivided whenever

$$\left\| h^k |\nabla_{k+1} \rho| \right\|_c > \delta C$$

for a fixed threshold δC.

This is the second most important refinement.

4. Finally, we can subdivide cells according to standard error estimates for the solution to the Poisson Equation. This is the least important refinement and may be done in the crudest fashion. In the current programs, this step is entirely skipped.

During the self-consistent field iteration, ρ only changes little. Consequently, refinement needs only be done rarely, in the order of once every 15 SCF steps. At the same time as the refinement, coarsening may be performed according to the converse conditions.

Figures 5.3, 5.4, and 5.5 illustrate the three non-standard refinement steps for a large, one-dimensional cell containing two electronic charges somewhere. The example is the same as that given in Figure 5.2, that is, $\rho(\mathbf{x}_q) = 0$ for all the quadrature points $\mathbf{x}_q$. In Figures 5.4, and 5.5, the linear approximations to $\rho(\mathbf{x})$ are only meant to illustrate the convergence: The real linear FE-representations do not simply connect the quadrature points, but are computed using projection operators. However, despite this slight inaccuracy, the arising issues are the same.

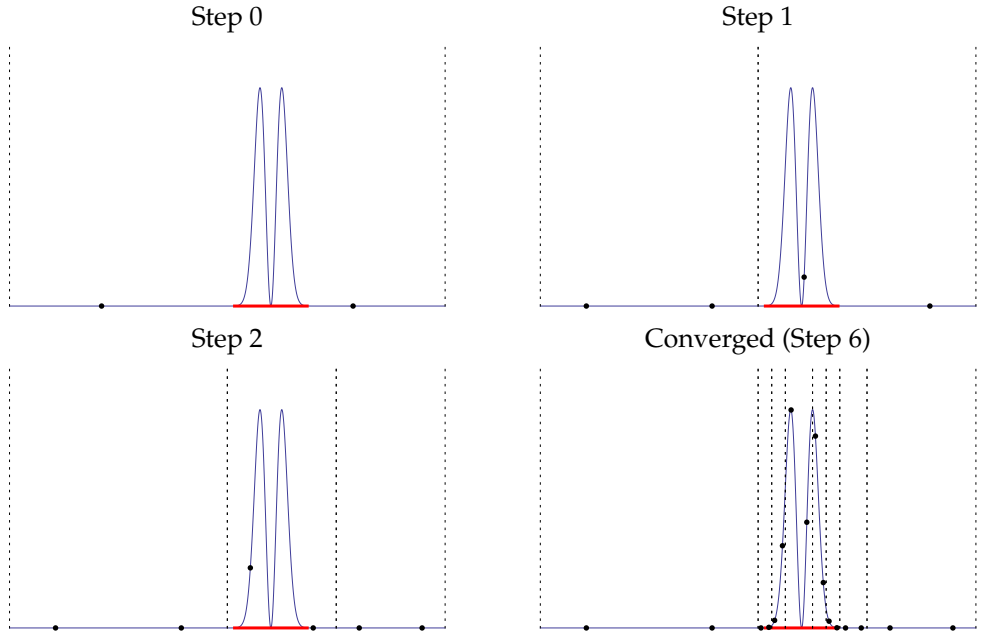


Figure 5.3: Initial refinement by centers and cutoff-radius shown in one dimension. Two electrons are somewhere in a cell with two Gaussian quadrature points, corresponding to linear elements. The cell is recursively subdivided according to step 1 and Equation (5.7.3). After the final refinement, we are able to approximate density integrals by quadrature. Dotted lines signify cell boundaries, and the dots show the quadrature points.

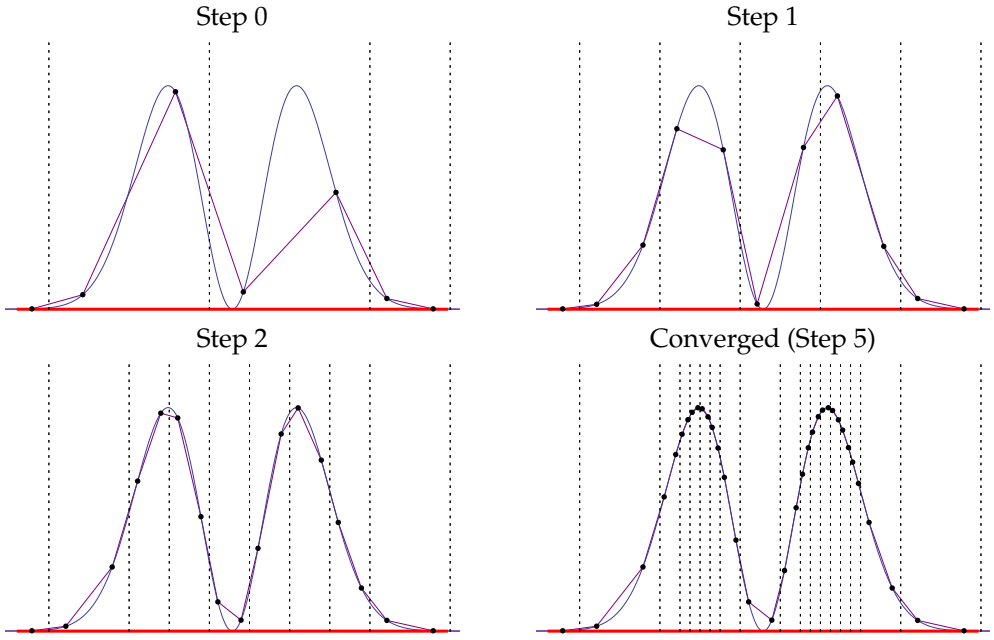


Figure 5.4: Refinement by density with $\delta e = 0.2$ electron per cell. We have zoomed in on the support of $\rho(\mathbf{x})$, since the rest of the mesh is untouched. Notice that regions with high curvature, but low electron content, yield poor point-wise approximations to $\rho(\mathbf{x})$. This is rectified in the next pass.

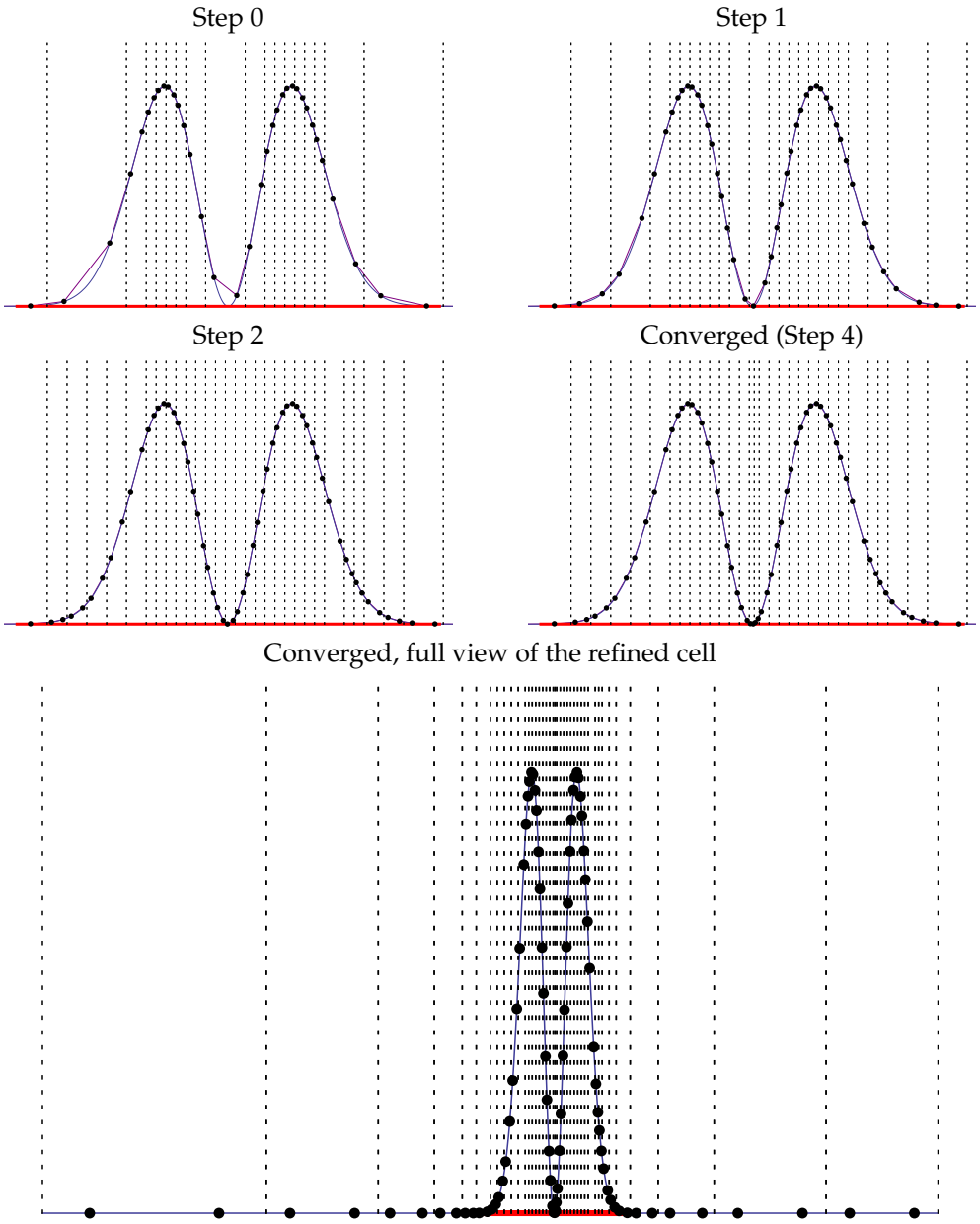


Figure 5.5: Refinement by ρ-curvature $\left\| h^k |\nabla_{k+1}\rho| \right\|_c$ with $\delta C = 2$. Since the elements are linear, we refine when $\left\| h \left| \frac{\partial^2 \rho}{\partial x^2} \right| \right\|_c > \delta C$. In the final step, a regularization pass has been made to ensure $\frac{1}{2} \leq \frac{\text{diam}(c_{i+1})}{\text{diam}(c_i)} \leq 2$ for neighbouring cells.

CHAPTER 6

Software

6.1 Introduction

This chapter provides a rudimentary overview of the functionality of the software that has been developed. Unfortunately, space limitations prohibit giving a thorough description. More comprehensive documentation will be written and distributed with the software when it matures to a point where it is ready for public release.

6.1.1 libspace

libspace is a space discretization library that implements the theory described in Chapter 2 as well as functionality needed to support the computational method described in Chapter 5. It abstracts away the details of Sobolev space and finite element methods with the intent of allowing the same application code to run with various finite element or finite difference backends.

A snapshot of the source code can be downloaded at

http://sturmian.kvante.org/libspace/libspace.tar.gz

or the latest version can be checked out from source control at

git@github.com:jamesavery/libspace.git

6.1.2 qscf

The linear scaling SIESTA-like DFT code qscf has been extended to use the finite element method with methods described in Chapter 5 by way of the libspace library. For copyright reasons, qscf can not be downloaded, but the software can be run from Copenhagen University's systems at Danish Center for Scientific Computing (DCSC), and the source code can be perused at

fend03.dcsc.ku.dk:/home/avery/work/qscf+fem/

The executable program is located at /home/avery/Linux-x86_64/bin/openmp-qscf. The FEM+DFT-code is found in the subdirectory qscf+fem/src/femorb/. A login must be obtained from Copen-

hagen University. All input data, programs for automatically generating environments and experiments, as well as scripts for analyzing and visualizing output data, can be found at

`fend03.dcsc.ku.dk:/home/avery/work/qscf-calculations/SET`

Many more experiments than the fraction described in this thesis are found there. The results can additionally be downloaded as Mathematica data files from

`http://sturmian.kvante.org/SET/results.tar.gz`

At present, the only fully supported finite element backend is the `deal.II` library (Bangerth et al.). A very rudimentary finite difference backend is included, as is a partially implemented but not yet functioning finite element backend using `libmesh` (Kirk et al. [2006]).

6.2 `libspace` library overview

The main interface to the library is the `Discretization` class, which provides basic Ritz-Galerkin functionality for approximate function spaces over $\mathbb{R}^d$. It is parameterized with the types F_h, a class describing discretized functions, and O_h, a class describing discretized operators. Figure 6.1 shows the class template inheritance hierarchy. It exports the following operations:

- Constructing finite representations of continuous functions: $C(\mathbb{R}^d) \ni f \mapsto f_h \in F_h$.
- Integrating:

$$\begin{aligned} f_h &\mapsto \int d^d x \, f_h(\mathbf{x}) \\ (f_h, g_h) &\mapsto \langle f_h | g_h \rangle \\ (f_h, V_h, g_h) &\mapsto \langle f_h | V_h | g_h \rangle \end{aligned}$$

- Computing overlap and Laplacian matrices.
- Solving Poisson's Equation: $\rho_h \mapsto \phi_h$, where ϕ solves $-4\pi\epsilon(\mathbf{x})\nabla^2\phi(\mathbf{x}) = \rho(\mathbf{x})$.

The most important discretization interface is the `FESpace` class, which abstracts finite element methods over a domain $\Omega \subseteq \mathbb{R}^d$. It is a subclass of the `Discretization` class for both `PointFunctions` and `FEFunctions`, i.e. providing the basic operations for both discretization types. An `FESpace` class exports methods for

- Loading and constructing meshes.
- Exporting meshes and FE-functions.
- The adaptive refinement methods described in Chapter 5.
- Setting material ids for DOFs and spatial regions.
- Setting regions with fixed solutions.
- Setting dielectric regions.
- Setting Dirichlet and von Neumann boundary value conditions.
- Converting between `PointFunction` and `FEFunction` representations.

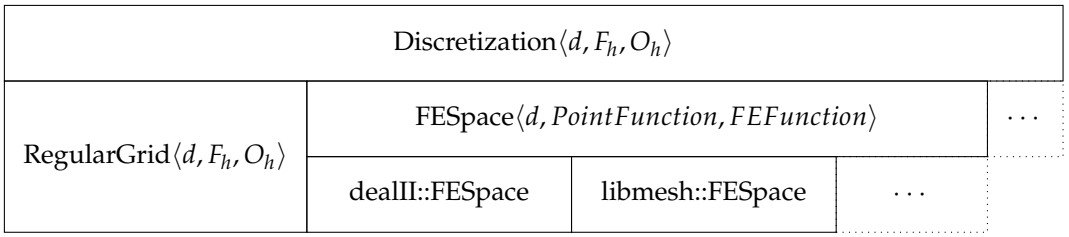


Figure 6.1: Crude overview of the discretization class hierarchy.

- Pointwise evaluation/interpolation.
- Differentiation.
- Estimating error.
- Controlling quadrature order, etc..

All these methods are accessed in a way that is agnostic towards the particular finite element implementation. While the only complete backend is done using the `deal.II`-library, a second backend using `libmesh` (Kirk et al. [2006]) is partially done, and a third using the `FeNiCS` framework (Logg and Wells [2010]) is planned. The various well-established finite element libraries each have their strengths and weaknesses, and the freedom to choose among them is useful. In addition, the data structures are given as template parameters to the `Discretization` and `FESpace` classes. The details of the data structures and storage methods are hidden behind an abstract container interface, and no code depends on them. For some applications, represention by simple dense STL vectors or flat floating point arrays may be preferable, both for the sake of code size, speed and so as not to depend on large libraries like `Trilinos`. In other situations, problem size dictates the use of distributed sparse arrays and matrices. This scheme allows flexibility with respect to the underlying representation.

CHAPTER 7

Modeling a single molecule, single electron transistor

7.1 Introduction

Figure 7.1 shows schematically the physical arrangement of a molecular single electron transistor (SET). A molecule rests on a dielectric. On each side of the molecule are a source and a drain, which might typically be gold electrodes. On the opposite side of the dielectric, a gate voltage V_g is applied, and between the source and drain electrodes another voltage $V_{sd} = V_s - V_d$ is applied.

Electron transfer occurs as tunneling events. It is important to distinguish between the Coherent Tunneling Regime and the Sequential Tunneling Regime. In the Coherent Tunneling Regime the molecule is in strong contact with both the source and the drain. In that case, when an electron passes from the source and across the molecule to the drain, the transfer takes place coherently. The tail of the broadened LUMO (Lowest Unoccupied Molecular Orbital) can allow passage of the the electron, if it is below the Fermi level of the electrons in the source and above the Fermi level of the electrons in the sink. Broadening is caused both by thermal vibrations proportional to $k_B T$, and by the finite lifetime τ_c of the coherent state, which produces a broadening $\Gamma \gg 1/\tau_c$ due to Heisenberg's uncertainty principle.

By contrast, in the weakly coupled Sequential Tunneling Regime, the tunneling events take place one at a time, the intervening time between them being enough for coherence to be lost. Thus, single electron transfers take place as discrete events, separated from each other by sufficient

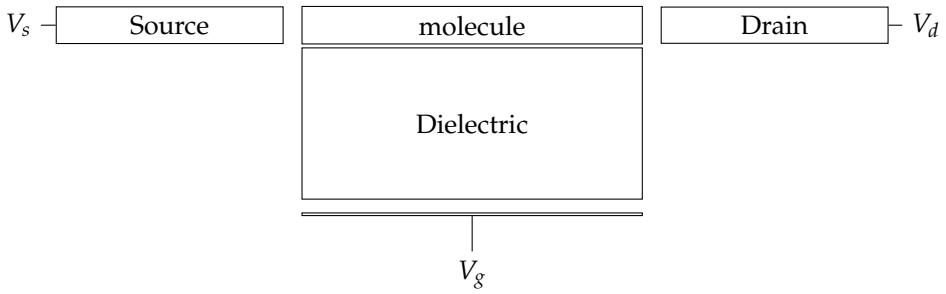


Figure 7.1: This diagram shows schematically the physical arrangement of a molecular single electron transistor (SET).

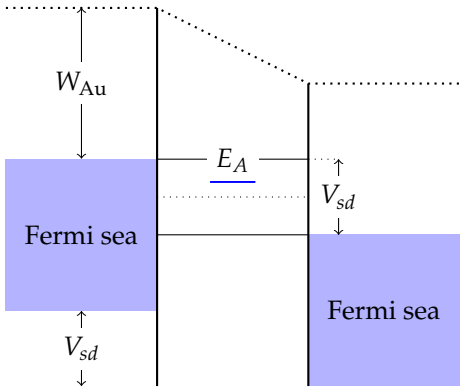


Figure 7.2: This figure shows schematically the potential energy as a function of distance for a molecular single electron transistor in the Sequential Tunneling Regeme. The applied source-drain potential V_{sd} has shifted the surface level of the Fermi sea in the source to a higher energy than that of the Fermi surface level in the drain. Meanwhile, the applied gate potential V_g has brought an electron affinity level of the molecule (EA) to an energy lower than the Fermi level of the source, allowing tunneling from the source to the molecule to take place.

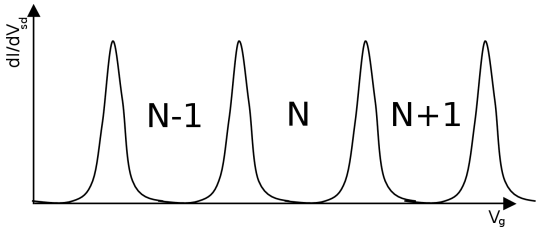


Figure 7.3: Schematic representation of the differential conductance for a single electron transistor. If the peaks are well separated, electron transfer occur only at these points. Between the peaks, the molecule is in a stable redox state and will neither accept nor donate electrons.

time for phase memory to be lost. The current work concentrates on molecular single electron transistors that operate in the Sequential Tunneling Regime.

The barriers that ensure weak coupling between the molecule and the source-drain electrodes can take the form either of small empty spaces, or chemical blocking groups. If the source-molecule gap and drain-molecule gap are large enough to prevent coherent current flow, but small enough to allow electron tunneling, then the SET is in the Sequential Tunneling Regime, as is illustrated in Figure 7.2. However, controlling the exact placement of the molecule to the needed accuracy is difficult, which makes chemical insulation preferable from an experimental standpoint. In the calculations in this chapter, as well as for the experiment by Kubatkin et al. described in Chapter 8, tert-butyl groups terminating the large organic molecule provide the blocking barrier.

In the Sequential Tunneling Regime, an electron can tunnel from the source to the molecule, provided that an electron affinity (EA) of the molecule lies at a potential less than the Fermi surface of the source. The position of the electron affinity level can be adjusted by changing the gate potential V_g. The extra electron remains on the molecule for a long enough period for coherence to be lost. A second tunneling event can take the electron to the drain. Alternatively the gate potential V_g can be changed so as to bring the next electron affinity level into the appropriate range and another extra electron can be gained by the molecule. The process can be continued until several extra electrons have been gained by the molecule. With other values of V_g, the molecule can lose electrons, so that it becomes positively charged.

From Figure 7.3 we can see, that in order for a system to operate as a transistor, the peaks of in differential conductance must be well separated. If the peaks are smeared out and overlap, then current can flow unhindered at all gate voltages. The condition necessary for the peaks to be well separated is

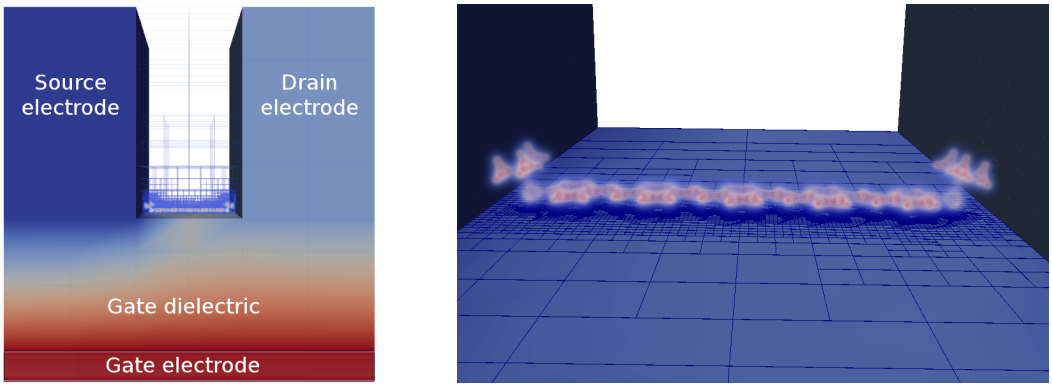
$$E_c, \Delta \gg k_B T, \Gamma \tag{7.1.1}$$

where E_c is the charging energy, and Δ denotes the level spacings. The typical level spacings Δ and charging energies E_c for other types of SETs, for example quantum dots, are in the order of magnitude $0.1 - 1$ meV . This means that they are smaller or on the order of magnitude of Boltzmann's constant times the temperature unless the temperature is lower than a few Kelvin. By contrast, the molecular SETs can potentially function at room temperature, or even higher. The SET, which is treated in this chapter, uses a tert-butyl terminated OPV5 molecule, and has charging energies and level spacings on the order of an electron volt. This more than satisfies Equation (7.1.1): At room temperature, $k_B T = 0.026$eV, and at $100^\circ C$, $k_B T = 0.032$eV. In general, the larger the system, the more closely spaced are the energy levels. Thus, a small system, like a single molecule, has the advantage of energy levels that are widely separated compared to $k_B T$.

7.2 The SET environment being modeled

The molecule is situated on top of a dielectric of thickness H between two large electrodes, with the lowest lying nucleus at distance d_y from the surface, the leftmost nucleus at distance d_x from the source electrode and the rightmost nucleus similarly at distance d_x from the drain. When not otherwise specified, the calculations below will use the following dimensions:

$$d_y = 1\text{Å}, \quad d_x = 1\text{Å}, \quad H = 50\text{Å} \tag{7.2.1}$$



(a) (b)

Figure 7.4: (a) OPV5-tBu molecule in a single electron transistor environment, coloured by effective potential $V^{eff}(\mathbf{x})$. Here shown with $V_g = 3V$ and $V_{sd} = 1V$. (b) Close-up view of the molecule, showing the electron density $\rho(\mathbf{x})$. Note the extremely high level of detail in the region where $\rho(\mathbf{x})$ is large.

The gate electrode is modeled by simply applying a Dirichlet boundary with value V_g on the downwards face of the dielectric. The source and drain electrodes are applied Dirichlet conditions with values $V_{sd}/2$ and $-V_{sd}/2$. The source and drain electrodes must, however, remain as part of the space, since any electron density overlapping with the electrodes must be included in the energy integrals. The boundaries that are not part of electrodes are simply given homogeneous von Neumann conditions; when the dimensions of the system are sufficiently large, this approximates the open boundary condition.

7.3 Calculations in vacuum

To check the correctness of the basic FEM+LDA code, we can compare to calculations done using standard quantum chemical software. These will not allow us to include effects from an external environment, which must be tested separately, but we can check against our calculations in vacuum.

7.4 Validity of our method; Energies of OPV5-tBu in vacuum

While we are not aware of other software that can treat molecules the size of OPV5-tBu in a large electrostatic environment, we can obtain results for the *isolated* molecule using standard quantum chemical software. This allows us to assess the quality of our calculations: both as a check for errors, and to determine whether LDA is accurate enough to treat this molecule. Reference calculations were made using `Gaussian '03` Frisch et al. [2003], using the polarized double-zeta basis set, `cc-pVDZ`, and the B^{3}LYP exchange-correlation functional.

Figure 7.5 and Table 7.1 shows qualitative agreement between the `Gaussian` B^{3}LYP results and our FEM+LDA calculations. To the extent that the OPV5-tBu molecule can be properly described by B^{3}LYP, our finite element LDA achieves this as well.

		B^{3}LYP	FEM+LDA[1]
ΔE^{-5}	(5^{th} EA)	-7.50	-7.75
ΔE^{-4}	(4^{th} EA)	-5.24	-5.24
ΔE^{-3}	(3^{rd} EA)	-2.83	-3.12
ΔE^{-2}	(2^{nd} EA)	-0.78	-0.644
ΔE^{-1}	(1^{st} EA)	1.60	1.56
ΔE^{0}	(1^{st} IP)	6.01	5.55
ΔE^{+1}	(2^{nd} IP)	8.43	7.71
ΔE^{+2}	(3^{rd} IP)	10.3	9.96
ΔE^{+3}	(4^{th} IP)	12.8	12.2
ΔE^{+4}	(5^{th} IP)	13.7	14.3

Table 7.1: Charging energies $\Delta E^q = E^{q+1} - E^q$ in eV for the isolated OPV5-tBu molecule. B^{3}LYP-values were computed using `Gaussian '03`, as is discussed in the text.

7.5 Charging energies and addition energies

Table 7.2 shows total energies relative to the neutral molecule as well as the charging energies $\Delta E^q = E^{q+1} - E^q$, i.e. the vertical ionization energies and electron affinities. Values are computed for the isolated molecule in vacuum and for the molecule placed in the SET environment as described in the previous section. An experimental study (Papaefthimiou et al. [2002]) found the ionization energy of Ooct-OPV5 (which is expected to have similar electrical properties to OPV5-tBu) to be 5.45 ± 0.1eV, in approximate agreement with our results for OPV5-tBu. The 2% difference may be caused by the somewhat larger size of the Ooct-OPV5 molecule resulting in slightly smaller charging energies; however part of the difference may be due to calculational inaccuracy. A test calculation using our method for the isolated benzene molecule yielded the ionization energy to 1% of the experimental value given in the NIST CCCBDB database (9.15 eV versus 9.24 eV). In summary, we expect the calculated charging energies to be accurate to within a couple of percent.

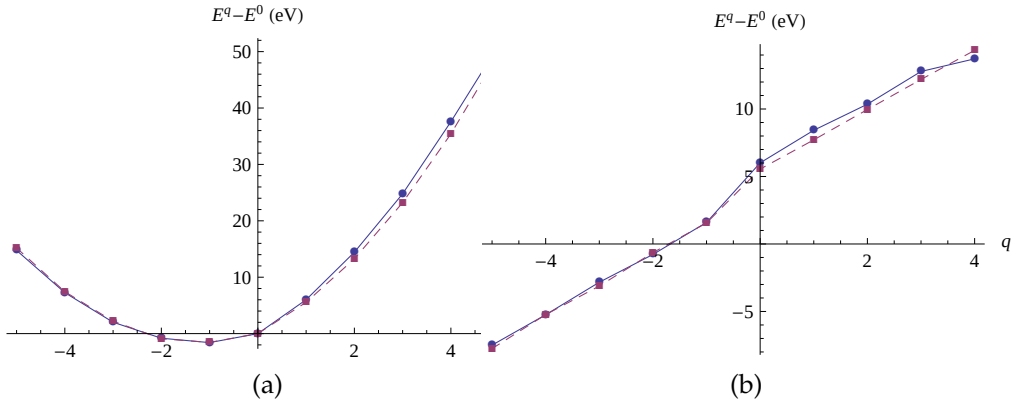


Figure 7.5: Comparison to `Gaussian '03` calculations of OPV5-tBu using B^3LYP. (a) Total energy difference from the neutral molecule as a function of the added charge q. (b) Vertical charging energies $\Delta E^q = E^{q+1} - E^q$. The solid lines are our FEM+LDA results, and the dashed lines are the results from `Gaussian '03`.

We note from Table 7.2 that the effect of placing the OPV5-tBu molecule in the SET environment is a reduction in the *differences* between charging energies. In fact, as we can see in Figure 7.6, this reduction is linear in the charge q with $\Delta E^q_{\text{SET}} \approx \Delta E^q_{\text{Vac}} - (1.48q + 0.39)$eV ($\pm 0.1$eV). This linear reduction of the charging energies corresponds to an approximately constant reduction of the molecular *addition energies* $\Delta^2 E^q = \Delta E^{q+1} - \Delta E^q$ of roughly 1.46eV. Table 7.3 and Figure 7.7 shows the addition energies of the molecule in the SET compared to the isolated molecule. In Section 7.7, we explain the approximately linear reduction of the charging energies, and the corresponding roughly constant reduction of $\Delta^2 E^q$.

		In vacuum	In SET
ΔE^{-5}	(5^{th} EA)	-7.75	-0.849
ΔE^{-4}	(4^{th} EA)	-5.24	0.293
ΔE^{-3}	(3^{rd} EA)	-3.12	0.951
ΔE^{-2}	(2^{nd} EA)	-0.644	1.97
ΔE^{-1}	(1^{st} EA)	1.56	2.70
ΔE^{0}	(1^{st} IP)	5.55	5.15
ΔE^{+1}	(2^{nd} IP)	7.71	5.83
ΔE^{+2}	(3^{rd} IP)	9.96	6.57
ΔE^{+3}	(4^{th} IP)	12.2	7.39
ΔE^{+4}	(5^{th} IP)	14.3	8.04

Table 7.2: Calculated charging energies (eV) for the OPV5-tBu molecule in the SET environment compared to the isolated molecule in vacuum.

q	-5	-4	-3	-2	-1	0	+1	+2	+3
$\Delta^2 E^q_{\text{Vac}}$	2.52	2.12	2.47	2.2	3.99	2.16	2.25	2.23	2.16
$\Delta^2 E^q_{\text{SET}}$	1.14	0.66	1.02	0.73	2.45	0.68	0.742	0.816	0.65

Table 7.3: Calculated charging energies (eV) in for the OPV5-tBu molecule in the SET environment compared to the isolated molecule in vacuum.

7.6 Redox transitions and charge stability diagrams

Figure 7.8 shows total energy curves of the charge states at zero source-drain voltage as a function of the ground voltage V_g. The energies are corrected for the work function for gold, chosen to be $W_{Au} = 5.28$eV (Rivière [1966]). The lowest line for a given V_g is thus the most stable charge state for that voltage when the molecule is in contact with a gold electron reservoir. The crossings between two lowest states is the voltage at which extra charge is added or removed.

Figure 7.9 shows the corresponding approximation to the full charge stability diagram, the colours signifying the number of charging energies that are within the bias window $[-V_{sd}/2; V_{sd}/2]$. Within the black diamonds, the molecule is stable, and no charge can be added or removed; within the coloured regions, electrons flow to or from the electrodes. That is, the transistor is open around the crossings, where a source-drain voltage results in electron transport.

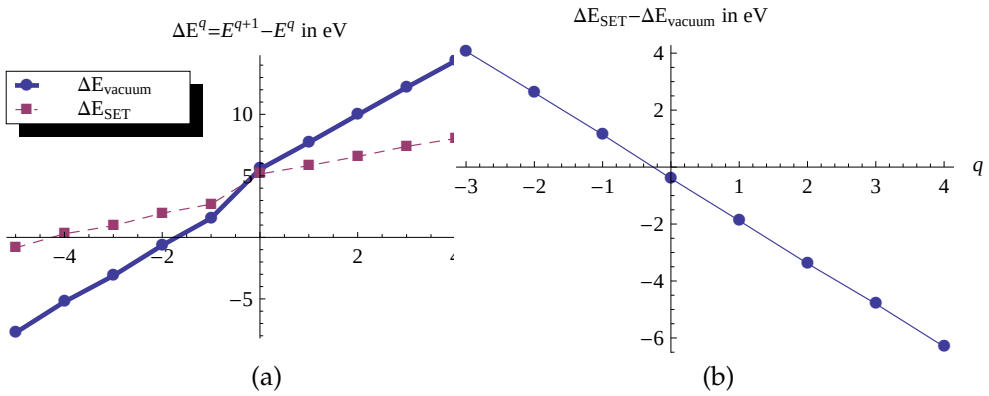


Figure 7.6: (a) Calculated charging energies for the OPV5-tBu molecule in the SET environment compared to the isolated molecule. (b) Reduction in charging energies from vacuum to the SET environment.

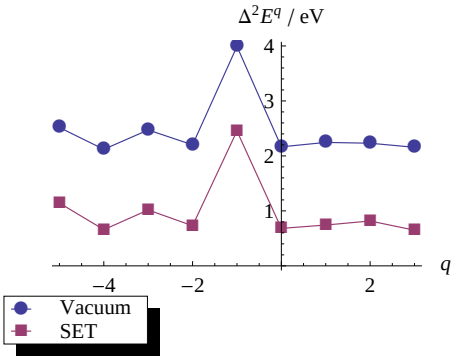


Figure 7.7: Calculated addition energies for the isolated OPV5-tBu molecule in vacuum, compared to the molecule in the SET environment.

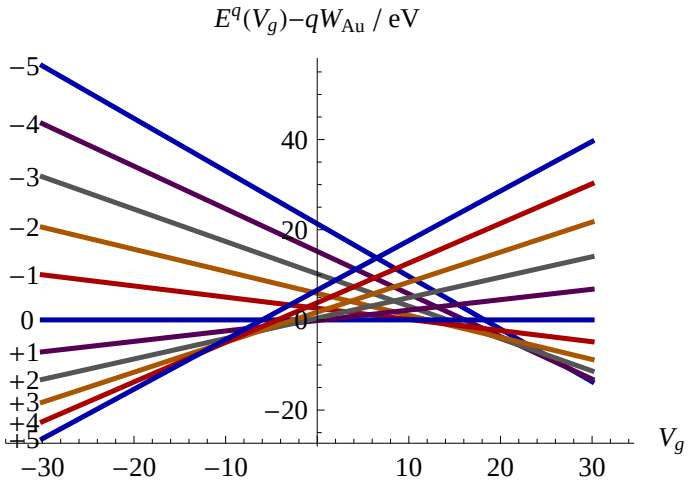
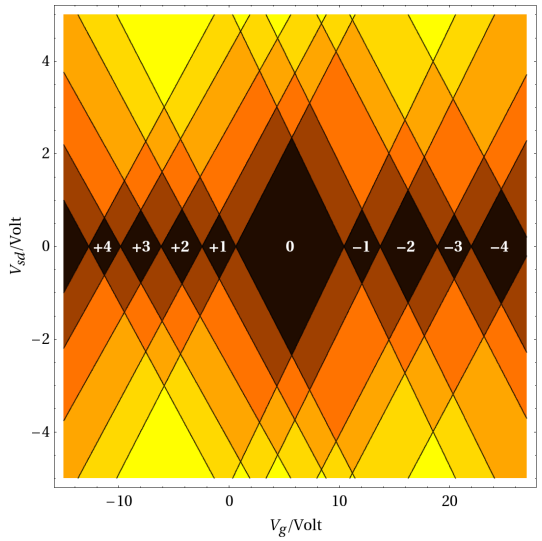


Figure 7.8: Total energies corrected for the workfunction of gold. The lines are labeled with the charge q. The energy of the neutral atom is chosen as the zero-point.



(a)

q	$V(q \leftrightarrow q+1)$	L_q	$\Delta^2 E^{q-1}_{\mathrm{SET}}$	α^q	$\tilde{\alpha}^q$
-4	22.0	5.50	1.142	0.24	0.22
-3	18.9	3.07	0.659	0.24	0.23
-2	13.7	5.18	1.017	0.24	0.23
-1	10.4	3.28	0.730	0.25	0.24
0	0.613	9.82	2.453	-	0.25
+1	-2.45	3.06	0.682	0.23	0.23
+2	-6.16	3.71	0.742	0.23	0.22
+3	-9.85	3.69	0.816	0.22	0.22

(b)

Figure 7.9: (a) Approximate charge stability diagram for OPV5-tBu in the SET-environment. The labels in the diamonds are added charge q; i.e., as the gate voltage V_g is increased, more electrons will move from the gold to the molecule. (b) Redox transition voltages, diamond widths, and gate coupling constant $\alpha = \frac{1}{q}\frac{\partial E}{\partial V_g}$. $\tilde{\alpha}$ is the approximation calculated from L_q and $\Delta^2 E$ by Equation (7.6.4).

7.6.1 How to read the diamond plots

We assume that the total energy is linearly dependent on the gate-potential V_g. This is the case for stretched out molecules like OPV5. For our calculations, this is true to within six digits. This yields:

$$E^q(V_g) = E^q(0) + q\alpha^q V_g \tag{7.6.1}$$

where $\alpha^q = \frac{1}{q}\frac{\partial E^q}{\partial V_g}$ is the gate coupling coefficient. When $V_{sd} = 0$, the transition between redox state q and $q+1$ occurs when the energy difference between the two redox states is equal to the work function of gold. This is attained for

$$V(q \leftrightarrow q+1) = -\frac{\Delta E^q(0) - \mathrm{W_{Au}}}{\alpha^q} \tag{7.6.2}$$

The width L_q in Volts of the diamond corresponding to the stable charge state q is then

$$\begin{aligned} L_q \equiv V(q-1 \leftrightarrow q) - V(q \leftrightarrow q+1) &= -\frac{\Delta E^q(0) - \mathrm{W_{Au}}}{\alpha^q} \\ &= -\frac{\Delta E^{q-1}(0) - \mathrm{W_{Au}}}{\alpha^{q-1}} + \frac{\Delta E^q(0) - \mathrm{W_{Au}}}{\alpha^q} \end{aligned} \tag{7.6.3}$$

Assuming $\alpha^q \approx \alpha^{q-1}$, we then obtain

$$L_q \approx \frac{\Delta E^q(0) - \Delta E^{q-1}}{\alpha^q} = \frac{\Delta^2 E^{q-1}}{\alpha^q} \tag{7.6.4}$$

Equation (7.6.4) is only roughly accurate, yielding diamond widths approximately within 5%. The redox transition points, and hence the gate voltages at which conductance can occur when applying a source-drain potential, are thus determined entirely by α and $\Delta E^q - \mathrm{W_{Au}}$, i.e. the difference between the charging energies and the work function, and the widths of the charge states are controlled by $\Delta^2 E^q$ and α.

7.6.2 Variable oxide thickness and distance from the dielectric

Figures 7.10 and 7.11 show, how IP, EA, the addition energy $U = \Delta^2 E^{q-1}$, and the gate coupling coefficient, $\alpha = \frac{1}{q}\frac{\partial E}{\partial V_g}$, change with the distance d_y of the molecule to the gate oxide. In addition, the dependence of α on the gate thickness H is shown. The energies at $V_g = 0$ are largely unaffected by the gate thickness, so we omit this plot. α, on the other hand, is approximately inversely proportional to H.

As we saw from Equation (7.6.4), the width of the charge state q in the charge stability diagram is approximately $L_q \approx \Delta^2 E^{q-1}(0)/\alpha$. If we calculate the width L_0 of the neutral charge state, we can see that it grows approximately linearly both in d_y and in H:

$$\begin{aligned} L_0 &\approx 0.205\tfrac{V}{\text{Å}} \cdot H && \text{when } d_y = 1\text{Å}, H \geq 30\text{Å} \\ L_0 &\approx 8V + 2.27\tfrac{V}{\text{Å}} \cdot d_y && \text{when } H = 50\text{Å} \end{aligned} \tag{7.6.5}$$

7.7 Analysis of the results

Understanding the energies in vacuum

We can understand the charging and addition energies by way of the Kohn-Sham spectrum of the OPV5-tBu molecule. While, due to electron correlation and self interaction, Koopmans' theorem Koopmans [1934] cannot be directly applied to density functional theory, analogous results

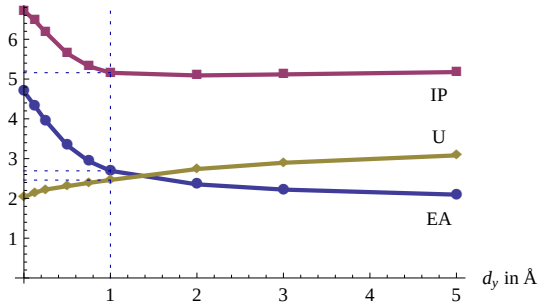


Figure 7.10: The first ionization energy IP, electron affinity EA, and addition energy U of OPV5-tBu as a function of the distance d_y to the gate oxide. The gate thickness is kept at a constant $H = 50$Å, and no external voltage is applied. The dotted lines show the values at $d_y = 1$Å, the distance used in our calculations.

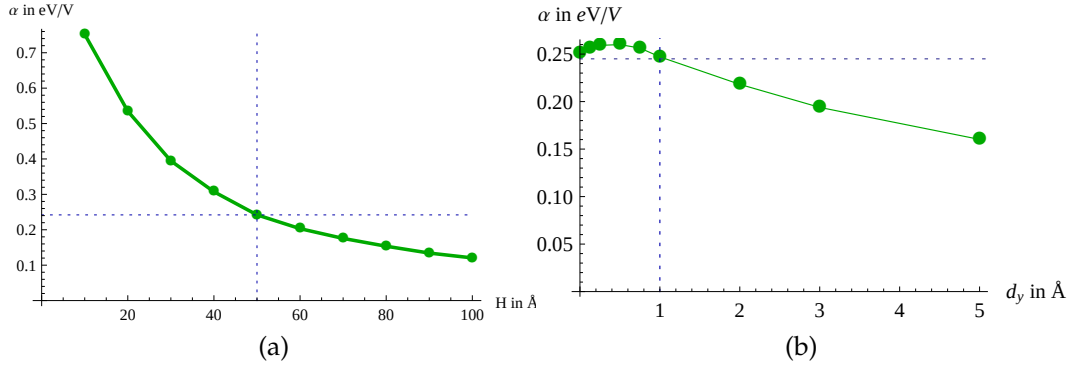


Figure 7.11: The gate coupling coefficient $\alpha = \frac{1}{q}\frac{\partial E^{q+1}}{\partial V_g}$ of the OPV5-tBu molecule (a) as a function of gate thickness H at $d_y = 1$Å, and (b) the distance d_y from gate to molecule.

do exist connecting the highest occupied Kohn-Sham eigenvalue to the IP Janak [1978], Perdew and Levy [1997], Politzer and Abu-Awwad [1998], Luo et al. [2006]. Specifically, Perdew and Levy showed that for the exact DFT solution, the vertical ionization energy $IP = -\varepsilon_N$; for the approximate DFT methods that are currently realizable, we have (Luo et al. [2006])

$$\Delta E = E^{N-1} - E^N \approx -\varepsilon_N^N + K[\rho^N] \tag{7.7.1}$$

where N is the total number of electrons, and the functional-dependent term K contains electrostatic and self-interaction terms. For the charge states of an extended molecule, K is nearly constant. For OPV5-tBu, we have found $K \approx -1$eV. The addition energies are then:

$$\Delta^2 E^N = \Delta E^{N-1} - \Delta E^N \approx \varepsilon_{N-1}^{N-1} - \varepsilon_N^N \tag{7.7.2}$$

Figure 7.12 shows Equation (7.7.2) to be in excellent agreement with the results.

We can further relate the charging and addition energies directly to the spectrum of (one-electron) Kohn-Sham eigenvalues of the *neutral* molecule in vacuum. The Kohn-Sham energies for the different charging states of OPV5-tBu are approximately linear in q with $\varepsilon_i^q \approx \varepsilon_i^0 - 2.1\text{eV} \cdot q$, whereby

$$\Delta E \approx -\varepsilon_i^q + K \approx -\varepsilon_i^0 + q \cdot 2.1\text{eV} - 1\text{eV} \tag{7.7.3}$$

and

$$\Delta^2 E \approx \varepsilon_{N-1}^0 - \varepsilon_N^0 + 2.1\text{eV} \tag{7.7.4}$$

That is, the addition energies $\Delta^2 E$ are well approximated simply by the Kohn-Sham spectrum shifted upwards by an additive constant corresponding to a quadratic contribution to the total energy due primarily to interelectron repulsion. We see from Figure 7.12 that Equation (7.7.4) reproduces the addition energies somewhat less accurately than Equation (7.7.2). However, the distinguishing features are preserved: the overall shape of $\Delta^2 E$ is closely matched to the spectrum of Kohn-Sham energies for the neutral molecule. As we will see below, quadratic energy contributions such as electrostatic effects merely shift $\Delta^2 E$ up or down by constant amounts, unless the dependence on q of the spatial distribution of charge is altered significantly, the molecular geometry changes, or another phenomenon with super-quadratic energy-contribution occurs.

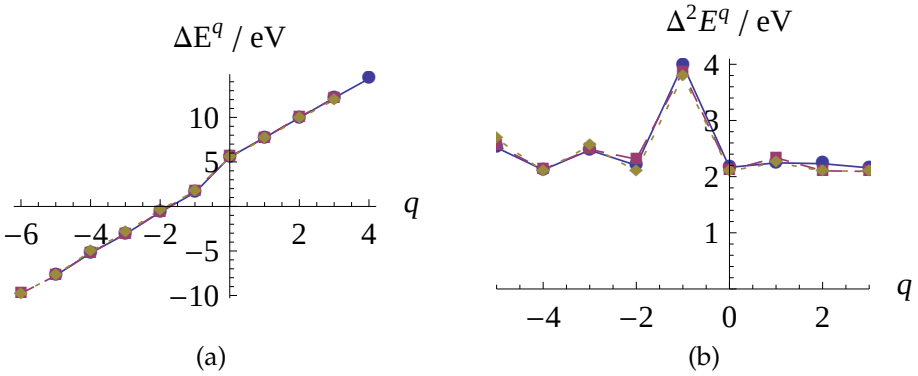


Figure 7.12: Addition energies for the isolated OPV5-tBu molecule compared to approximations from Kohn-Sham eigenvalues: —●— : ΔE^q_{Vac} (left) and $\Delta^2 E^q_{\text{Vac}}$ (right); - -■- - : Approximations (7.7.1) and (7.7.2); and ···◆··· : Approximations (7.7.3) and (7.7.4).

Understanding the energy shift from vacuum to SET

Figure 7.6(b) shows that the shift in charging energies ΔE^q from vacuum to the SET environment is linear in q, i.e. the change in $\Delta^2 E^q$ is (mostly) constant. This can be explained using simple classical electrostatics: The molecule is stretched out, mostly flat, and has a large surface area lying closely against the gate dielectric; only the extreme points of the molecule are close to the source/drain-electrodes. At $V_{sd} = 0$, we consequently expect the dominant effect from the SET on the molecule to be the electrostatic interaction with the dielectric. We can then roughly approximate the shift by neglecting the effect of the electrodes and treating the problem with an induced image-charge distribution:

$$\begin{aligned} E^q_{\text{SET}} &\approx E^q_{\text{Vac}} + \frac{1}{2}\left(\frac{\epsilon-1}{\epsilon+1}\right) \iint \delta\rho(\mathbf{x}_1)\frac{1}{r_{12}}\delta\rho_{Img}(\mathbf{x}_2) \\ &= E^q_{\text{Vac}} - \frac{1}{2}\left(\frac{\epsilon-1}{\epsilon+1}\right) \iint \delta\rho(\mathbf{x}_1)\frac{1}{r_{12}}\delta\rho(\mathbf{x}_2 - (0,0,2d_y)) \end{aligned}$$

where $\delta\rho \equiv \rho^q - \rho^0$. Assume now that the spatial distribution of added charge $\Delta\rho^q(\mathbf{x}) = \rho^{q+1}(\mathbf{x}) - \rho^q(\mathbf{x})$ is fixed, i.e. $\Delta\rho^{q+1} \approx \Delta\rho^q$. Then we can factor $-q^2$ out from the integral and obtain

$$\begin{aligned} E^q_{\text{SET}} &\approx E^q_{\text{Vac}} - q^2 C \\ \Delta E^q_{\text{SET}} &\approx \Delta E^q_{\text{Vac}} - (2q+1)C \\ \Delta^2 E^q_{\text{SET}} &\approx \Delta^2 E^q_{\text{Vac}} - 2C \end{aligned} \tag{7.7.5}$$

where the coefficient C is defined as $C = \frac{\epsilon-1}{2(\epsilon+1)} I(\Delta\rho, d_y)$, with I being the integral on the right hand side of Equation (7.7). C depends only on the constant distribution of charge difference

$\Delta\rho(\mathbf{x})$, the dielectric constant ϵ, and the distance d_y from molecule to dielectric. Note, that while we have only considered the contribution from the dielectric, if we were to include the electrodes or added other classical electrostatic element, it would not change Equation (7.7.5) qualitatively. So long that we assume a fixed $\Delta\rho$, only the coefficient C is affected. In this case, the contribution to the total energy must be $\propto q^2$, it must be $\propto (2q+1)$ to the charging energies ΔE^q, and constant to the addition energies $\Delta^2 E^q$.

We can give a quantitative estimate of the validity of our calculations in the SET environment with the following rough approximation: Assume a charge difference that is equally distributed over a two-dimensional rectangle A. Then

$$C(A) = \frac{9}{22|A|^2} \int_A \frac{1}{(x_1 - x_2)^2 + (z_1 - z_2)^2 + 4d_y^2} \tag{7.7.6}$$

and the shift in ionization energies is

$$\Delta E^q_{\text{SET}} - \Delta E^q_{\text{Vac}} \approx -(2q+1)C(A) \tag{7.7.7}$$

By setting the rectangle to the xz-bounding box of the OPV5-tBu molecule (38.24Å $\times$ 5.45Å), this approximation yields

$$\Delta E^q_{\text{SET}} - \Delta E^q_{\text{Vac}} \approx -0.740(2q+1)\text{eV} \tag{7.7.8}$$

which is to be compared to the value obtained from the full calculation

$$\Delta E^q_{\text{SET}} - \Delta E^q_{\text{Vac}} = -0.738(2q+0.528) \tag{7.7.9}$$

The small difference in the coefficient C is in part due to the crude approximation of assuming a uniform rectangular charge distribution, in part due to inclusion of the entire SET-environment; However, as is apparent from the discussion above, the difference in the intersection with the q-axis ($2q = -0.53$ as opposed to $2q = -1$) must be due to charge localization that varies with q, since otherwise the factor $(2q+1)$ must appear. Similar calculations for the more compact benzene molecule yields a crossing at $2q = -0.8$, and for the hydrogen ions close to $2q = -1$, substantiating this claim.

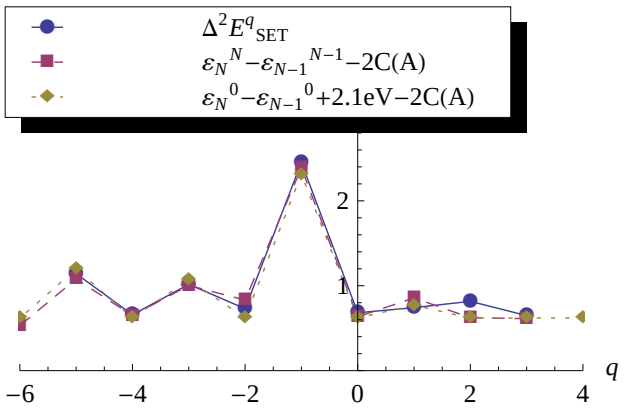


Figure 7.13: Addition energies for OPV5-tBu in the SET environment compared to approximations calculated from Equation (7.7.6) and Equation (7.7.2) and (7.7.4), respectively.

CHAPTER 8

Comparison to experiments by Kubatkin et al.

In their 2003 Nature article Kubatkin et al. [2003], Kubatkin and coworkers describe one of the first realized single molecule single electron transistors using an OPV5-tBu molecule as the capacitor. Their SET was constructed as follows: Kubatkin and co-workers prepared a planar gate electrode made of aluminium metal covered with aluminium oxide ($\approx$ 5 nm thick) on a chip of oxidized silicon. They used standard electron beam lithography to define a shadow mask, which was used to deposit the gold lead electrodes. They then introduced the chip into a vacuum chamber immersed in liquid helium. Afterwards, during a single vacuum cycle, they performed the following operations: They deposited two gold electrodes through a shadow mask by condensing gold vapor onto the substrate held at 4.2 K. By using an oblique evaporation angle together with in situ conductance measurements, they were able to fine-tune the tunneling gap between the two electrodes to a few nanometers. By reducing the tilt angle for the oblique evaporation they were able to decrease the source-drain gap. They changed the tilt stepwise, narrowing the gap between the leads by 5 nm at each step. At each step they deposited a 2-nm-thick film of gold through the mask, and the sample was checked for non-zero tunneling conductance. Eventually they fabricated two self-aligned and self-sharpening gold electrodes with a tunneling gap of a few MΩ between them. By annealing the sample up to 100 K they increased the gap width to a few GΩ, which corresponds to a tunneling gap width of roughly 2 nm.

8.1 Reported results

Table 8.1 lists the results reported by Kubatkin et al., and Figure 8.2 shows the measured addition energies graphically. There are two remarkable features of these numbers, as we can see on Figure 8.2:

i. The peak representing the addition energy $\Delta^2 E^{-1} = \text{IP} - \text{EA}$ has disappeared in the SET environment.

ii. Both measurements and calculations in the solution, and the measurements in the SET environment yield very small values of $\Delta^2 E^q$.

Why are these two features so remarkable? First, we recall from Section 7.7 that any energy contribution that is up to quadratic in q (such as simple electrostatic effects) must only shift $\Delta^2 E^q$ up or down by a constant amount. To reduce the IP-EA gap $\Delta^2 E^{-1}$ to almost zero, while at the

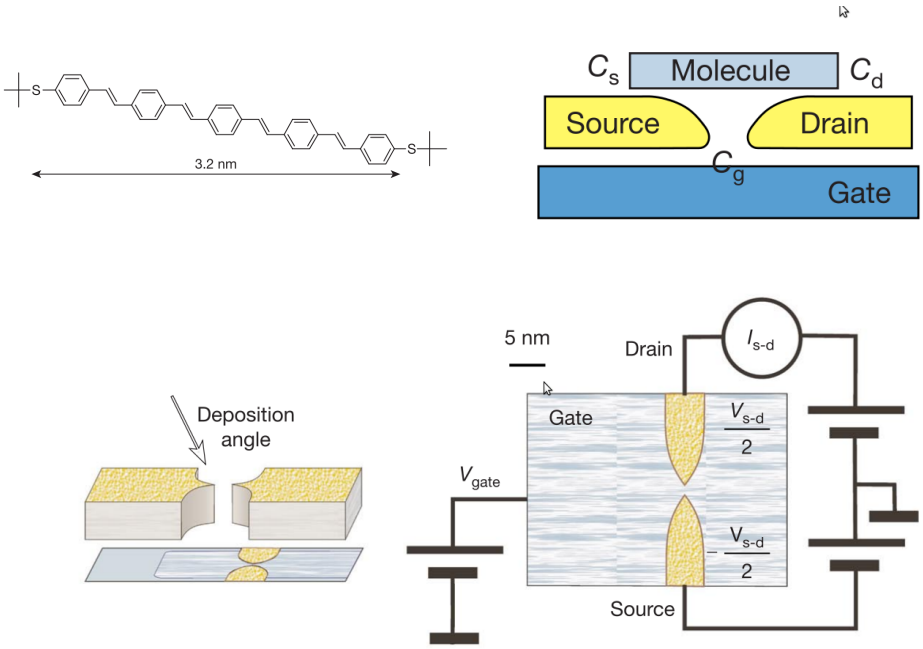


Figure 8.1: Schematic overview of experimental setup (Figure 1 of Kubatkin et al. [2003])

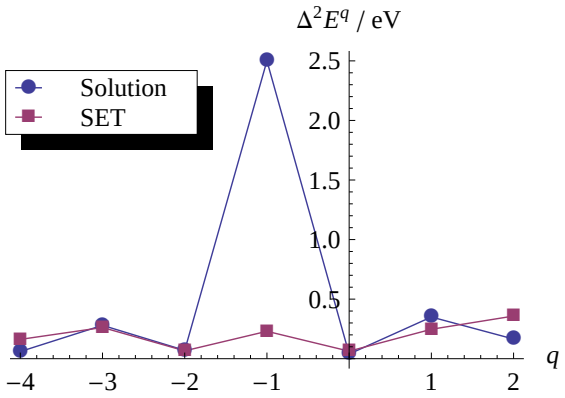


Figure 8.2: Plot of the addition energies shown in Table 8.1. The values in the solution are comprised of estimates from semiempirical AM1 calculations (right of peak), values obtained from electrochemical midway potentials Heinze et al. [1987] (left of peak), and estimate from the absorption edge of OPV5 in the solution. The SET-values are all experimental and derived from the charge stability diagram of the constructed single electron transistor.

same time leaving all the other addition energies essentially unchanged, requires a large super-quadratic and likely non-polynomial effect. Secondly, the small values correspond to total energy functions with very low curvature. In the solution, the numbers predict a total energy curve that is nearly piecewise linear in the given region, with a sharp break at $q = 0$. However, for the SET environment, the numbers yield a total energy curve that is nearly linear everywhere! If the addition energies measured by Kubatkin et al. are correct, then both of these are very interesting results that suggest complex effects at play. In Subsection 8.3, we will see the dramatic effect of Point (ii) on the autoionization properties of the OPV5-tBu molecule.

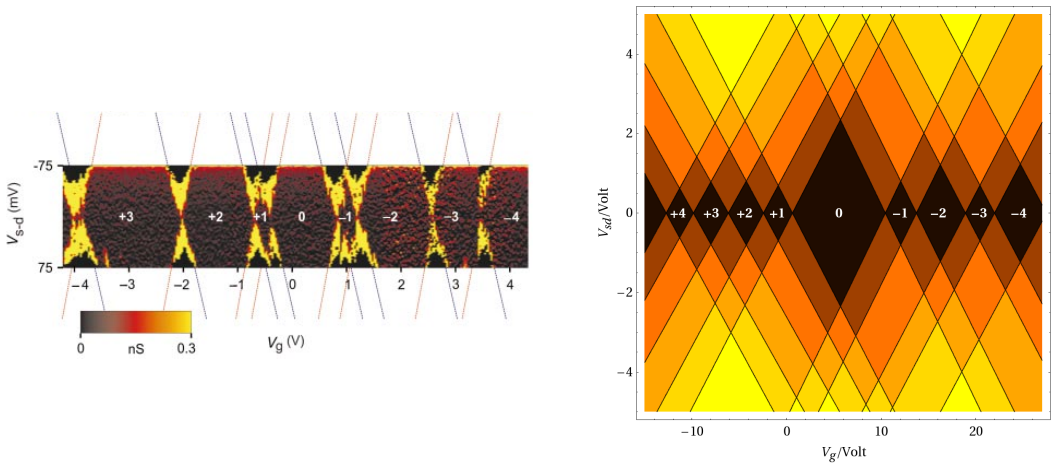
Table 1 **Addition energies and redox states**

Redox state	n	V_g (n) (V)	$I(n \rightarrow n+1)^{SET}$ (meV)			$I(n \rightarrow n+1)^{other}$ (meV)	
−4	4	3.53*	†	‡	*		
−3	3	2.64			162	60§	
−2	2	1.20			262	280§	
−1	1	0.84		60	66	70§	66#
0	0	−0.42		312	230	2,500 ‖	230#
1	−1	−0.77	70	56	63	39¶	66#
2	−2	−2.13	390	266	247	350¶	
3	−3	−4.10			358	165¶	

Symbols are defined in the text.
*Data in these two columns are from Fig. 2a, representing the essence of many measurements on the same device/molecule.
††Data in these two columns are from two additional devices prepared independently. The ratio $C_g/(C_s + C_d + C_g) = 0.2$.
§From electrochemical midway potentials (ref. 20) multiplied by a scaling factor of 0.5.
‖Estimated from absorption edge of OPV5 in solution situated at 450 nm (own data).
¶Estimated from computed ionization potentials (AM1, dielectric constant = 2.5, as suggested from data in ref. 25) multiplied by a scaling factor of 0.16.
#Addition energies calculated by a Hubbard model (see Methods) representing the molecule as 38 p_z carbon orbitals with an on-site repulsion U and a transfer integral t between connected atoms. Image charges have been included by changing the on-site energy on one or both terminal benzene ring(s) by an amount equal to the mirror image charge energy, Φ, given by $\Phi_{image} = 27.2\ (\text{eV Å})/(\varepsilon_r\ a)$, where ε_r is the relative dielectric constant and a the distance between the charge and its image in ångströms. The values listed in Table 1 are obtained by fitting t to the optical HOMO–LUMO gap of the pristine OPV5 molecule (2,500 meV). The two independent parameters $(\varepsilon_r\ a)$ and U are obtained by fitting the model to the experimental addition energies obtained for $n = -1, 0, +1$. The resulting values are: $t = 3.92$ eV, $U = 4.27$ eV, $(\varepsilon_r\ a) = 4.7$ Å. The value of $(\varepsilon_r\ a)$ is in good agreement with the expected distance between the plane of the benzene ring and its image when the benzene ring is in van der Waals contact with the electrode. The charges are thus expected to be separated by roughly twice the van der Waals radius of sp^2-hybridized carbon (~3 Å). U is also well within the expected range for on-site repulsion in organic molecules.

Table 8.1: Table 1 in Kubatkin et al. [2003]. The addition energies $\Delta^2 E^{q-1} = \Delta^2 E^{n+1}$ are listed as $I(n \rightarrow n+1)$, and the redox transition points $V(q \leftrightarrow q+1)$ are listed under $V(n)$.

8.2 Charge stability plots



(a) (b)

Figure 8.3: (a) Figure 2(a) in Kubatkin et al. [2003]. (b) FEM+LDA-calculations in the SET environment with $V_{sd} = 0$.

8.2.1 Comparing the charge stabilities

The values of $V(q \leftrightarrow q+1)$ and $\Delta^2 E^q_{\text{SET}}$ are read off from Table 8.1. From these, the widths L_q of the stable charge states and the gate coupling coefficient $\alpha = \frac{1}{q}\frac{\partial E}{\partial V_g}$ are computed according to Section 7.6.1. Table 8.2 shows a comparison to the calculated values. Unsurprisingly, due to the great large discrepancy in $\Delta^2 E^q_{\text{SET}}$, the calculated charge states are much wider that the experimentally determined ones. However, the gate coupling coefficient is comparable. This indicates that the molecule is situated within a few Ångström of the dielectric: According to our calculations, with an oxide thickness of 50Å , the gate coupling coefficient $\alpha \approx 0.18\text{eV}/V$ corresponds to a distance of approximately 4Å (see Figure 7.11).

		q	-3	-2	-1	0	+1	+2	+3
$V(q \leftrightarrow q+1)$	Kubatkin et al.	(V)	2.64	1.20	0.84	-0.42	-0.77	-2.13	-4.10
$V(q \leftrightarrow q+1)$	Calculated		18.9	13.7	10.4	0.613	-2.45	-6.16	-9.85
L_q	Kubatkin et al.	(V)	0.89	1.44	0.36	1.26	0.35	1.36	1.97
L_q	Calculated		3.07	5.18	3.28	9.82	3.06	3.71	3.69
$\Delta^2 E^{q-1}_{\text{SET}}$	Kubatkin et al.	(eV)	0.162	0.262	0.066	0.23	0.063	0.247	0.358
$\Delta^2 E^{q-1}_{\text{SET}}$	Calculated		0.659	1.017	0.730	2.453	0.682	0.742	0.816
α^q	Kubatkin approx. (7.6.4)	(eV/V)	0.18	0.18	0.18	0.18	0.18	0.18	0.18
α^q	Calculated $\frac{1}{q}\cdot\frac{\partial E^q_{\text{SET}}}{\partial V_g}$		0.24	0.24	0.25	-	0.23	0.23	0.22
α^q	Calculated, approx. (7.6.4)		0.23	0.23	0.24	0.25	0.23	0.22	0.22

Table 8.2: Comparison of charge stability dimensions.

In the following section, we turn to the task of investigating the large discrepancy between the calculated results and the experiment.

8.3 Autoionization

Due to the reduction in addition energies in the electrostatic environment, the autoionization properties of the OPV5-tBu molecule are altered when the molecule is deposited in the SET even when no external voltages are applied. Reducing the addition energies corresponds to reducing the curvature of the total energy as a function of q, as was seen on Figure 7.7. Consequently, the energy minimum changes position. The minimum of the total energy curve, however, does not directly yield the spontaneous level of ionization: We must take into account the energy needed to remove an electron from the environment. The most stable charge state has charge

$$q_0 = \operatorname*{argmin}_q \left(E^q - qW\right) \tag{8.3.1}$$

where W is the work function for the environment surrounding the molecule. That is, the molecule will spontaneously release or obtain electrons until it reaches the charge state q_0.

In the SET environment, we can assume that electrons will come from the gold electrodes, and let W be the work function for gold; A good value is $\text{W}_{\text{Au}} = 5.28\text{eV}$ recommended by Rivière [1966]. However, depending on temperature, purity, surface properties and multiple other factors, the work function for gold can vary up to a few electron volts. In order to predict the range of autoionization one will find in experiments, it is then necessary to allow for a variance in W. To assess the level of autoionization found in Kubatkin et al., we further need to allow for a variance in the ionization energy: Whatever causes the large difference in addition energies for the Kubatkin et al. setup may also cause the ionization energy to be quite changed. Since no charging energies were reported, we must leave IP as a free parameter in the following.

Before we continue, we note that we can reconstruct the charging energies and total energy curves from addition energies using the relation

$$\sum_{q=a}^{b-1} \Delta f_q = f_b - f_a \tag{8.3.2}$$

That is, given a single value ΔE^{q_0}, we can derive all other values ΔE^q from the $\Delta^2 E^q$'s, et cetera. Equation (8.3.2) yields:

$$\begin{aligned} E^q &= E^0 + \sum_{n=0}^{q-1} \Delta E^n \\ &= E^0 + \sum_{n=0}^{q-1} \left(\Delta E^0 + \sum_{m=0}^{n-1} \Delta^2 E^m \right) \\ &= E^0 + q\Delta E^0 + \sum_{n=0}^{q-1} \sum_{m=0}^{n-1} \Delta^2 E^m \end{aligned} \tag{8.3.3}$$

In other words, we can write the work function corrected energy as

$$E^q - qW = E^0 + \sum_{n=0}^{q-1} \sum_{m=0}^{n-1} \Delta^2 E^m + q(\mathrm{IP} - W) \tag{8.3.4}$$

Consequently, we do not need to vary W and the ionization energy separately; Varying the parameter IP $-$ W suffices.

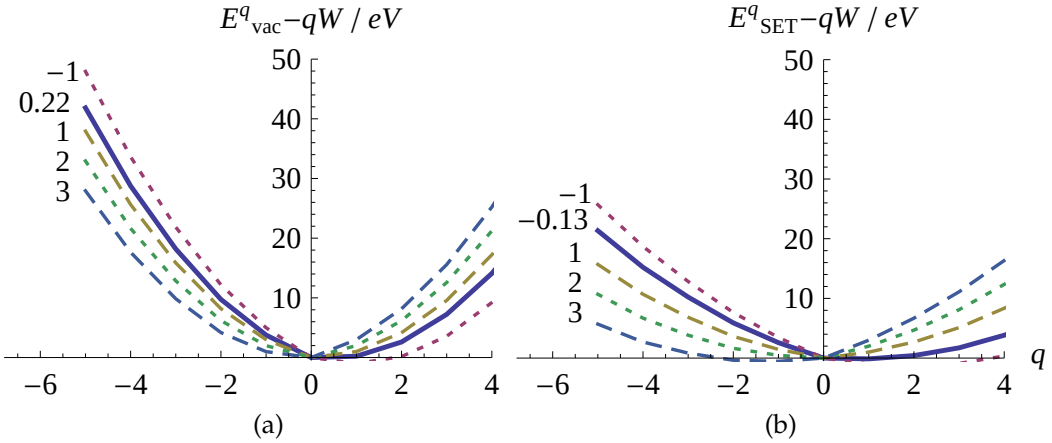


Figure 8.4: Total energy curve with work function correction (a) in vacuum, and (b) in the SET. The energies were derived from the calculated $\Delta^2 E^q$ given the specified values of $IP - W$ (given in eV). The minima of the curves, which are listed in Table 8.3, predict the level of autoionization. The solid curve corresponds to our calculated IP and $W = 5.28$eV.

Figure 8.4 shows the work function corrected energy curves corresponding to the range $IP - W \in [-1\mathrm{eV}; 3\mathrm{eV}]$ for our FEM+LDA calculations. The minima, corresponding to the autoionized charge states, are given in Table 8.3. We observe that while the lowered addition energies in the SET environment do make the molecule more volatile, it is mostly stable both in vacuum and deposited on the SET. Within the most common range of W_{Au}, it is predicted to either remain in its neutral form or donate a single electron. Assuming the calculated IP is correct, a work function of around 2.5eV for the environment is required for the OPV5-tBu to spontaneously acquire an electron, while the work function must be around 6eV for the molecule to donate a single electron.

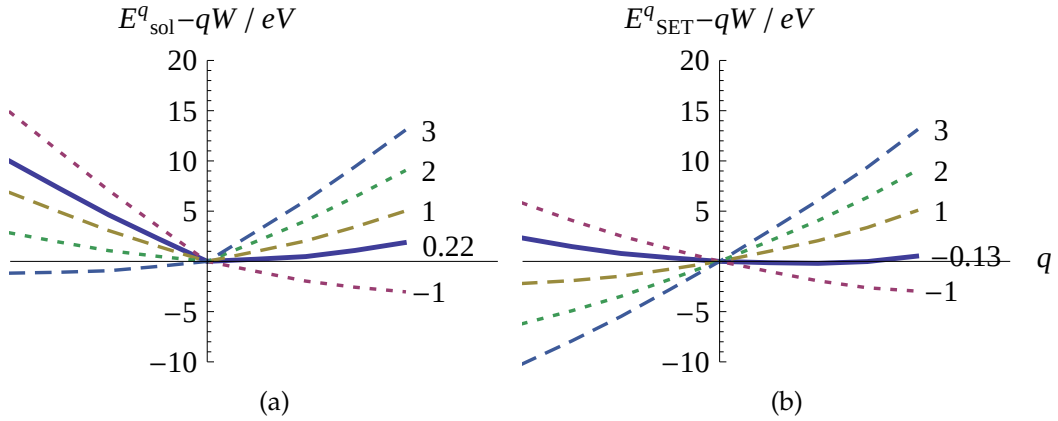


Figure 8.5: Total energies with work function correction derived from the addition energies reported by Kubatkin et al. (a) reported for OPV5-tBu in a solution, and (b) measured for OPV5-tBu in the SET environment. The values for $IP - W$ are the same as above. Note the change of scale compared to Figure 8.4.

Calculated with FEM+LDA						
IP$-W$ in eV		-1	0	1	2	3
Stable charge	in vacuum	+1	0	0	0	0
state q_0	in SET	+2	0	0	0	-1

Kubatkin et al.						
IP$-W$ in eV		-1	0	1	2	3
Stable charge	in solution	+5	0	0	0	-4
state q_0	in SET	+6	0	-5	-10	-16

Table 8.3: Autoionization of the OPV5-tBu molecule derived from addition energies for a range of IP $- W$, where W is the work function for the environment. The lower table is derived for from the addition energies reported by Kubatkin et al. in Kubatkin et al. [2003]. For charges lower than $q = -4$, values for $\Delta^2 E^q$ were not available; In this case, a quadratic polynomial was fitted to the total energies for $q \leq 0$, the minimum of which predicts the level of ionization.

The situation in the experimental setup is in quite strong contrast to this: As previously noted, the shallow addition energies measured and calculated by Kubatkin et al. yield a very low curvature of the total energy as a function of q. Consequently, small changes in the absolute value of the work function or ionization energy lead to drastically different autoionization properties. This is especially true in the SET environment, where there is no peak at $\Delta^2 E^{-1} = IP - EA$. As can be seen from Figure 8.5 and Table 8.3, the measured addition energies predict the OPV5-tBu molecule to be moderately stable in the solution, but wildly volatile when placed in the SET environment: The numbers predict that the molecule spontaneously gives up 6 electrons at IP $- W = -1$eV, but acquires 5 electrons at IP $- W = 1$eV, a change of 11 electrons due to a difference in IP $- W$ of just 2eV, a variation that is not unreasonable for experimental setups. If the measured addition energies are correct, it is thus possible that a completely different part of the spectrum of OPV5-tBu was measured than that around the neutral charge state. To determine *which* part of the spectrum this is, it is necessary to determine actual ionization energy of OPV5-tBu in the experimental SET environment, as well as the work function.

8.3.1 Discussion

In summary, there are a number of possible reasons that calculations did not reproduce the effect observed by Kubatkin et al.

1. Theoretically, it is possible that strong charge localization can cause the highly nonlinear reduction in addition energies. However, for electrostatic effects to almost exactly cancel the first addition energy $U = \Delta^2 E^{-1}$ while at the same time keeping all the other addition energies constant, quite elaborate conditions must hold. We have attempted a number of rough calculations taking such an effect into account, and have found it very difficult to reproduce anything close to this behaviour.

2. As the previous section showed, the extremely low addition energies measured by Kubatkin et al. correspond to a very volatile molecule that easily gains or relinquishes many electrons. The missing peak for the first addition energy may then be due to having observed a separate part of the spectrum.

3. It is not only possible, but likely, that the environment in the experimental setup introduces a change of molecular conformation, causing the drastic change of spectrum. Our calculations use a fixed geometry optimized for OPV5-tBu in vacuum, and thus would not reproduce such an effect.

Part II

Efficient multi-center electron repulsion integral computation for exponential type orbitals

CHAPTER 9

Introduction

One of the first achievements of quantum theory was the exact solution of the Schrödinger equation for a hydrogen-like atom, i.e. a single electron moving in the attractive potential of a single nucleus. The solutions to this problem are of the form

$$\chi_{nlm}(\mathbf{x}) = R_{nl}(r)Y_{lm}(\hat{\mathbf{x}}) \tag{9.0.1}$$

where $R_{nl}(r)$ is a polynomial times a decaying exponential in r, and Y_{lm} is a spherical harmonic. We call the solution to a one-electron Schrödinger equation an *orbital*. For many-electron atoms, exact solutions have until now not been obtained, because the many-electron Schrödinger equation is not separable. However, approximate solutions can be obtained, built up from antisymmetrized products of one-electron orbitals similar to hydrogen-like orbitals.

In molecules, atomic wave functions are only slightly modified by their environments. The atomic core remains much the same as for an isolated atom, with only high-energy orbitals participating in bonding between atoms. Therefore, in treating molecules, it would be natural to use the same types of basis functions that have been so successful in treating atoms. Exponential type orbitals (ETO's), as used in atomic calculations, have the correct behaviour at the nuclei, and the correct asymptotic behaviour. However, a practical difficulty arose in using ETO's in molecular calculations: The mathematics needed for evaluation of many-center interelectron repulsion integrals was prohibitively difficult. Therefore, McWeeny, Boys, and others introduced orbitals based on Cartesian Gaussians, which are functions of the form

$$x^{n_x}y^{n_y}z^{n_y}e^{-\zeta r^2} \tag{9.0.2}$$

The *Gaussian expansion theorem* states that the product of two Cartesian Gaussians at different positions is a single Cartesian Gaussian located at an intermediate position, multiplied by a tensor in x, y, and z. This property makes the evaluation of interelectron repulsion integrals possible, since two-center densities are thereby converted into sums of single center densities.

The computational quantum theory based on Gaussians has been highly developed during the last 60 years, with many advances towards efficiency. However, Gaussian basis functions are intrinsically problematic because very many are needed in order to approximate the cusp at atomic nuclei, and because they decay much too rapidly at large distances. The cusps at the nuclei are exactly represented by ETO's. If the exponent is correctly chosen, ETO's also match the asymptotic behaviour. Therefore, during the last 60 years, while most computational quantum chemistry followed the path of Gaussian technology, there have been a number of efforts to overcome the barrier of efficiently evaluating many-center interelectron repulsion integrals using ETOs. Among the talented authors who have addressed this highly mathematical problem are Frank E. Harris,

H. H. Michels, E. O. Steinborn, E. J. Weniger, Charles A. Weatherford, H. Jones, F. Rico, J. Ramírez, R. Lópes, and Philip Hoggan. While there has been some success in tackling the problem, existing computer programs are still so slow as to be discouraging.

In the following sections, we propose several new methods for approaching the difficult problem of evaluating many-center, interelectron repulsion integrals. We introduce molecular orbitals that are linear combinations of atomic orbitals of a particular type called Coulomb Sturmians. Members of a Coulomb Sturmian basis set have the form

$$p(kr)e^{-kr}Y_{lm}(\hat{\mathbf{x}}) \tag{9.0.3}$$

where the exponent k is the same for all members of the basis set, and the radial part is a pure function of $kr := s$. If k is replaced by Z/n, the Coulomb Sturmians become identical with hydrogen-like orbitals. However, keeping k constant for the entire basis set results in quite different properties. Among these is that for each positive value of k, Coulomb Sturmians are complete in the sense that they span $H^1(\mathbb{R}^3)$.

The two main methods proposed here are based on Fourier transforms. We are faced with integrals of the form

$$J = \int d^3x_1 \int d^3x_2 \, \rho_1(\mathbf{x}_1) \, \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|} \, \rho_2(\mathbf{x}_2) \tag{9.0.4}$$

where the densities ρ_1 and ρ_2 are given by $\rho_i(\mathbf{x}) = \chi^*_{\tau_a}(\mathbf{x} - \mathbf{X}_a)\chi_{\tau_b}(\mathbf{x} - \mathbf{X}_b)$. Following Harris and others, we introduce the representation

$$\frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|} = \frac{1}{2\pi^2} \int d^3p \, \frac{1}{p^2} e^{i\mathbf{p}\cdot(\mathbf{x}_1 - \mathbf{x}_2)} \tag{9.0.5}$$

Then J becomes

$$J = 4\pi \int d^3p \, \frac{1}{p^2} \rho_1^t(\mathbf{p})\rho_2^t(-\mathbf{p}) \tag{9.0.6}$$

where $\rho_i^t(\mathbf{p})$ are the Fourier transforms of the densities. As we shall see below, the Fourier transforms of Coulomb Sturmians can be expressed in terms of four dimensional hyperspherical harmonics. We will show that the properties of hyperspherical harmonics can be used to evaluate the interelectron repulsion integrals J. This can be done exactly and rapidly in the case of single center densities, i.e.

$$\begin{aligned} \rho_1(\mathbf{x}) &= \chi^*_{\mu_1}(\mathbf{x} - \mathbf{X}_1)\chi_{\mu_2}(\mathbf{x} - \mathbf{X}_1) \\ \rho_2(\mathbf{x}) &= \chi^*_{\mu_3}(\mathbf{x} - \mathbf{X}_2)\chi_{\mu_4}(\mathbf{x} - \mathbf{X}_2) \end{aligned} \tag{9.0.7}$$

with $\mathbf{X}_1 \neq \mathbf{X}_2$. For two-center densities, the hyperspherical method can also be used, but it involves an expansion of each two-center density.

CHAPTER 10

Mathematical tools

10.1 Coulomb Sturmian basis sets

10.1.1 One-electron Coulomb Sturmians

Because of their completeness properties, one-electron Sturmian basis sets have long been used in theoretical atomic physics. Their form is identical with that of the familiar hydrogenlike atomic orbitals, except that the factor Z/n is replaced by a constant k. The one-electron Coulomb Sturmians can be written as

$$\chi_{nlm}(\mathbf{x}) = R_{nl}(r)Y_{lm}(\theta,\phi) \tag{10.1.1}$$

where $n \geq 1$, $0 \leq l \leq n-1$, and $-l \leq m \leq l$. Y_{lm} is a spherical harmonic, and the radial function has the form

$$R_{nl}(r) = \mathcal{N}_{nl}(2kr)^l e^{-kr} F\left(l+1-n|2l+2|2kr\right) \tag{10.1.2}$$

Here

$$\mathcal{N}_{nl} = \frac{2k^{3/2}}{(2l+1)!}\sqrt{\frac{(l+n)!}{n(n-l-1)!}} \tag{10.1.3}$$

is a normalizing constant, while

$$F\left(a|b|x\right) := \sum_{k=0}^{\infty} \frac{a^{\bar{k}}}{k!b^{\bar{k}}}x^k = 1 + \frac{a}{b}x + \frac{a(a+1)}{2b(b+1)}x^2 + \cdots \tag{10.1.4}$$

is a confluent hypergeometric function. Coulomb Sturmian basis functions obey the following one-electron Schrödinger equation (in atomic units):

$$\left[-\frac{1}{2}\nabla^2 - \frac{nk}{r} + \frac{1}{2}k^2\right]\chi_{nlm}(\mathbf{x}) = 0 \tag{10.1.5}$$

which is just the Schrödinger equation for an electron in a hydrogenlike atom with the replacement $Z/n \to k$. The functions in a Coulomb Sturmian basis set can be shown to obey and obey a potential-weighted orthonormality relation of the form:

$$\int d^3x\, \chi^*_{n'l'm'}(\mathbf{x})\frac{1}{r}\chi_{nlm}(\mathbf{x}) = \frac{k}{n}\delta_{n'n}\delta_{l'l}\delta_{m'm} \tag{10.1.6}$$

All of the functions in a such a basis set correspond to the same energy,

$$\epsilon = -\frac{1}{2}k^2 \tag{10.1.7}$$

In other words the basis set is isoenergetic. In the wave equation obeyed by the Sturmians, (10.1.5), the potential is weighted differently for members of the basis set corresponding to different values of n.

10.1.2 The Fock Projection

The Fock Projection maps momentum space onto the surface of a 4-dimensional sphere. Using this projection, V. Fock Fock [1935] was able to establish a simple one to one correspondence between Coulomb Sturmian basis functions and 4-dimensional hyperspherical harmonics. Fock's mapping explained the puzzling n^2-fold degeneracy of the hydrogen-like orbitals, and it is of great use for the practical evaluation of integrals involving Coulomb Sturmians. It allows one to draw on the rich literature describing properties of hyperspherical harmonics, and to make use of the hyperangular integration methods described in Section 10.3 below.

Coulomb Sturmian basis functions and their Fourier transforms are related by

$$\chi_{nlm}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^3}} \int d^3x \, e^{i\mathbf{p}\cdot\mathbf{x}} \chi_{nlm}(\mathbf{p}) \tag{10.1.8}$$

and by the inverse transform

$$\chi^t_{nlm}(\mathbf{p}) = \frac{1}{\sqrt{(2\pi)^3}} \int d^3x \, e^{-i\mathbf{p}\cdot\mathbf{x}} \chi_{nlm}(\mathbf{x}) \tag{10.1.9}$$

By projecting momentum-space onto the surface of a 4-dimensional hypersphere, V. Fock was able to show that the Fourier-transformed Coulomb Sturmians can be very simply expressed in terms of 4-dimensional hyperspherical harmonics through the relationship

$$\chi^t_{n,l,m}(\mathbf{p}) = M(p) Y_{n-1,l,m}(\mathbf{u}) \tag{10.1.10}$$

where

$$M(p) := \frac{4k^{5/2}}{(k^2+p^2)^2} \tag{10.1.11}$$

and

$$\begin{aligned} u_1 &= \frac{2kp_1}{k^2+p^2} \\ u_2 &= \frac{2kp_2}{k^2+p^2} \\ u_3 &= \frac{2kp_3}{k^2+p^2} \\ u_4 &= \frac{k^2-p^2}{k^2+p^2} \end{aligned} \tag{10.1.12}$$

The 4-dimensional hyperspherical harmonics are given by

$$Y_{\lambda,l,m}(\mathbf{u}) = \mathcal{N}_{\lambda,l} C^{1+l}_{\lambda-l}(u_4) Y_{l,m}(u_1,u_2,u_3) \tag{10.1.13}$$

where $Y_{l,m}$ is a spherical harmonic of of the familiar type, while

$$\mathcal{N}_{\lambda,l} = (-1)^\lambda i^l (2l)!! \sqrt{\frac{2(\lambda+1)(\lambda-l)!}{\pi(\lambda+l+1)!}} \tag{10.1.14}$$

is a normalizing constant, and

$$C_\lambda^\alpha(u_4) = \sum_{t=0}^{[\lambda/2]} \frac{(-1)^t \Gamma(\lambda+\alpha-t)}{t!(\lambda-2t)!\Gamma(\alpha)} (2u_4)^{\lambda-2t} \tag{10.1.15}$$

is a Gegenbauer polynomial. The first few 4-dimensional hyperspherical harmonics are:

$$\begin{aligned} Y_{0,0,0}(\mathbf{u}) &= \frac{1}{\sqrt{2}\pi} \\ Y_{1,0,0}(\mathbf{u}) &= \frac{-2u_4}{\sqrt{2}\pi} \\ Y_{1,1,0}(\mathbf{u}) &= \frac{-2iu_3}{\sqrt{2}\pi} \\ \vdots \quad & \vdots \qquad \vdots \end{aligned} \tag{10.1.16}$$

The relationships between hyperspherical harmonics, harmonic polynomials, and harmonic projection will be discussed in Section 10.3.

10.2 Gaussians

The use of Gaussians to evaluate many center interelectron repulsion integrals in quantum chemistry rests on the Gaussian product theorem, which allows two-electron densities to be expressed as finite sums of one-electron densities. The theorem can be written as follows:

Theorem 10.2.1 (Gaussian product theorem)**.** *Given two Gaussians centered on the points $\mathbf{X}_a$ and $\mathbf{X}_b$, their product yields a single Gaussian centered at an intermediate point $\mathbf{X}_c$ as follows:*

$$\begin{aligned} f(\mathbf{x}) &= e^{-a|\mathbf{x}-\mathbf{X}_a|^2} e^{-b|\mathbf{x}-\mathbf{X}_b|^2} \\ &= e^{-\frac{ab}{a+b}|\mathbf{X}_a-\mathbf{X}_b|^2} e^{-(a+b)|\mathbf{x}-\mathbf{X}_c|^2} \\ &= Ce^{-(a+b)|\mathbf{x}-\mathbf{X}_c|^2} \end{aligned} \tag{10.2.1}$$

where the intermediate point $\mathbf{X}_c$, and the multiplicative constant C are given by

$$\mathbf{X}_c := \frac{a\mathbf{X}_a + b\mathbf{X}_b}{a+b} \quad \textit{and} \quad C := e^{-\frac{ab}{a+b}|\mathbf{X}_a-\mathbf{X}_b|^2} \tag{10.2.2}$$

To include the angular dependence of the orbitals, Cartesian Gaussians are used. These are of the form

$$(x-X_a)^{n_x}(y-Y_a)^{n_y}(z-Z_a)^{n_z} e^{-a|\mathbf{x}-\mathbf{X}_a|^2} \tag{10.2.3}$$

From the Gaussian product theorem, it follows that the product of two Cartesian Gaussians, centered at different points, can be expressed as a sum of Cartesian Gaussians centered at an intermediate point.

The Fourier transformed form of the Gaussian product theorem can be stated as follows:

Theorem 10.2.2 (Fourier transformed Gaussian product theorem)**.** *Let*

$$f(\mathbf{x}) = e^{-a|\mathbf{x}-\mathbf{X}_a|^2} e^{-b|\mathbf{x}-\mathbf{X}_b|^2} \tag{10.2.4}$$

be the product of two Gaussian basis functions located on different centers. Then its Fourier transform is given by

$$f^t(\mathbf{p}) = \frac{1}{(2\pi)^{3/2}} \int d^3x \, e^{i\mathbf{p}\cdot\mathbf{x}} f(\mathbf{x}) = \tilde{C} e^{-\alpha p^2 + i\mathbf{p}\cdot\mathbf{X}_c} \tag{10.2.5}$$

where

$$\tilde{C} := \frac{e^{-\frac{ab}{a+b}|\mathbf{X}_a - \mathbf{X}_b|^2}}{[2(a+b)]^{3/2}} \tag{10.2.6}$$

and

$$\alpha := \frac{1}{4(a+b)} \qquad \mathbf{X}_c := \frac{a\mathbf{X}_a + b\mathbf{X}_b}{a+b} \tag{10.2.7}$$

The Fourier transform of the more general Cartesian Gaussian basis function

$$F_{\mathbf{n}}(\mathbf{x}) = (x - X_a)^{n_1}(x - X_b)^{n_2}(y - Y_a)^{n_3}(y - Y_b)^{n_4}(z - Z_a)^{n_5}(z - Z_b)^{n_6} f(\mathbf{x}) \tag{10.2.8}$$

is given by

$$\begin{aligned} F^t_{\mathbf{n}}(\mathbf{p}) &= \left(\frac{1}{i}\frac{\partial}{\partial p_1} - X_a\right)^{n_1} \left(\frac{1}{i}\frac{\partial}{\partial p_1} - X_b\right)^{n_2} \\ &\times \left(\frac{1}{i}\frac{\partial}{\partial p_2} - Y_a\right)^{n_3} \left(\frac{1}{i}\frac{\partial}{\partial p_2} - Y_b\right)^{n_4} \\ &\times \left(\frac{1}{i}\frac{\partial}{\partial p_3} - Z_a\right)^{n_5} \left(\frac{1}{i}\frac{\partial}{\partial p_3} - Z_b\right)^{n_6} f^t(\mathbf{p}) \end{aligned} \tag{10.2.9}$$

We can express this more general Fourier transform in the form

$$F^t_{\mathbf{n}}(\mathbf{p}) = f^t(\mathbf{p}) P(\mathbf{n}) \tag{10.2.10}$$

where $f^t(\mathbf{p})$ is given in Equation (10.2.5), and where $P(\mathbf{n})$ is a polynomial. These results have been utilized and refined during the last 60 years, and have yielded rapid interelectron repulsion integral packages. However, Gaussian basis functions remain intrinsically inferior in accuracy to Exponential Type Orbitals, and very many are needed to obtain good representations of wave functions.

10.3 Homogeneous and harmonic polynomials; Angular and hyperangular integrations

This section discusses methods for evaluating angular and hyperangular integrals. These methods are extremely important because of their speed and convenience. The section will also discuss hyperspherical harmonics, which are important to the methods that will be used in Part III, because of their relationship to the Fourier transforms of Coulomb Sturmian basis functions. The contents of this section is based on Avery [1989], where the interested reader may find a treatment in greater depth.

Harmonic polynomials are defined as homogeneous polynomials that satisfy the generalized Laplace equation. It will be seen that hyperspherical harmonics are identical with harmonic polynomial, apart from a power of the hyperradius equal to the degree of the harmonic polynomial.

Three methods of angular and hyperangular integration will be presented. The first is closely related to the theory of harmonic polynomials. The second, which solves issues of resource consumption for the first method, depends on expressing the integrals both in Cartesian coordinates and in terms of the hyperradius and generalized solid angle. The third method is taken from

Avery and Ørmen [1980], and it is derived using the independence of the angular integrals on rotations of the coordinate system. We can then write them as linear combinations of scalar products, determined by combinatorial arguments. The coefficients in the linear combination are determined using the special properties of degenerate cases. Michels' later treatment in Michels [1981] proves the same properties using the Multinomial Theorem.

10.3.1 Homogeneous and harmonic polynomials

A *monomial of degree n* in d coordinates is a product of the form

$$m_n = x_1^{n_1} x_2^{n_2} x_3^{n_3} \cdots x_d^{n_d} \tag{10.3.1}$$

where the n_j's non-negative integers and where their sum is equal to n.

$$n_1 + n_2 + \cdots + n_d = n \tag{10.3.2}$$

For example, x_1^3, $x_1^2 x_2$ and $x_1 x_2 x_3$ are all monomials of degree 3. Since

$$\frac{\partial m_n}{\partial x_j} = n_j x_j^{-1} m_n \tag{10.3.3}$$

it follows that

$$\sum_{j=1}^{d} x_j \frac{\partial m_n}{\partial x_j} = n m_n \tag{10.3.4}$$

A homogeneous polynomial of degree n (which we will denote by the symbol f_n) is a series consisting of one or more monomials, all of which have degree n. For example, $f_3 = x_1^3 + x_1^2 x_2 - x_1 x_2 x_3$ is a homogeneous polynomial of degree 3. Since each of the monomials in such a series obeys (10.3.4), it follows that

$$\sum_{j=1}^{d} x_j \frac{\partial f_n}{\partial x_j} = n f_n \tag{10.3.5}$$

This simple relationship has very far-reaching consequences: If we now introduce the generalized Laplacian operator

$$\Delta := \sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2} \tag{10.3.6}$$

and the hyperradius defined by

$$r^2 := \sum_{j=1}^{d} x_j^2 \tag{10.3.7}$$

we can show (with a certain amount of effort!) that

$$\Delta \left(r^\beta f_\alpha \right) = \beta(\beta + d + 2\alpha - 2) r^{\beta-2} f_\alpha + r^\beta \Delta f_\alpha \tag{10.3.8}$$

where α and β are positive integers or zero, β being even. We next define a *harmonic polynomial of degree n* to be a homogeneous polynomial of degree n which also satisfies the generalized Laplace equation:

$$\Delta h_n = 0 \tag{10.3.9}$$

For example, $h_3 = x_1^2 x_2 - x_3^2 x_2 + x_1 x_2 x_3$ is a harmonic polynomial of degree 3. Combining (10.3.8) and (10.3.9) we obtain

$$\Delta \left(r^\beta h_\alpha \right) = \beta(\beta + d + 2\alpha - 2) r^{\beta-2} h_\alpha \tag{10.3.10}$$

10.3.2 The canonical decomposition of a homogeneous polynomial

Every homogeneous polynomial f_n can be decomposed into a sum of harmonic polynomials multiplied by powers of the hyperradius. This decomposition, which is called *the canonical decomposition of a homogeneous polynomial*, has the form:

$$f_n = h_n + r^2 h_{n-2} + r^4 h_{n-4} + \cdots \tag{10.3.11}$$

To see how the decomposition may be performed, we can act on both sides of equation (10.3.11) with the generalized Laplacian operator Δ. If we do this several times, making use of (10.3.10), we obtain:

$$\begin{aligned}
\Delta f_n &= 2(d+2n-4)h_{n-2} + 4(d+2n-6)r^2 h_{n-4} + \cdots \\
\Delta^2 f_n &= 8(d+2n-6)(d+2n-8)h_{n-4} + \cdots \\
\Delta^3 f_n &= 48(d-2n-8)(d-2n-10)(d-2n-12)h_{n-6} + \cdots
\end{aligned} \tag{10.3.12}$$

and in general

$$\Delta^\nu f_n = \sum_{k=\nu}^{\lfloor \frac{n}{2} \rfloor} \frac{(2k)!!}{(2k-2\nu)!!} \frac{(d+2n-2k-2)!!}{(d+2n-2k-2\nu-2)!!} r^{2k-2\nu} h_{n-2k} \tag{10.3.13}$$

where

$$j!! := \begin{cases} j(j-2)(j-4)\cdots 4 \times 2 & j = \text{even} \\ j(j-2)(j-4)\cdots 3 \times 1 & j = \text{odd} \end{cases} \tag{10.3.14}$$

An important special case occurs when $\nu = n/2$. In that case, (10.3.13) becomes

$$\Delta^{n/2} f_n = \frac{n!!(d+n-2)!!}{(d-2)!!} h_0 \tag{10.3.15}$$

or

$$h_0 = \frac{(d-2)!!}{n!!(d+n-2)!!} \Delta^{n/2} f_n \tag{10.3.16}$$

We will see below that this result leads to powerful angular and hyperangular integration theorems.

10.3.3 Harmonic projection

Equations (10.3.12) or (10.3.13) form a set of simultaneous equations that can be solved to yield expressions for the various harmonic polynomials that occur in the canonical decomposition of a homogeneous polynomial f_n. In this way we obtain the general result:

$$\begin{aligned}
h_{n-2\nu} &= \frac{(d+2n-4\nu-2)!!}{(2\nu)!!(d+2n-2\nu-2)!!} \\
&\times \sum_{j=0}^{\lfloor \frac{n}{2}-\nu \rfloor} \frac{(-1)^j (d+2n-4\nu-2j-4)!!}{(2j)!!(d+2n-4\nu-4)!!} r^{2j} \Delta^{j+\nu} f_n
\end{aligned} \tag{10.3.17}$$

If we let $n - 2\nu = \lambda$, this becomes

$$\begin{aligned} O_\lambda[f_n] = h_\lambda &= \frac{(d+2\lambda-2)!!}{(n-\lambda)!!(d+n+\lambda-2)!!} \\ &\times \sum_{j=0}^{\lfloor \lambda/2 \rfloor} \frac{(-1)^j (d+2\lambda-2j-4)!!}{(2j)!!(d+2\lambda-4)!!} r^{2j} \Delta^{j+\frac{1}{2}(n-\lambda)} f_n \end{aligned} \tag{10.3.18}$$

Here O_λ can be thought of as a projection operator that projects out the harmonic polynomial of degree λ from the canonical decomposition of the homogeneous polynomial f_n. The projection is of course taken to be zero if λ and n have different parities.

10.3.4 Generalized angular momentum

The generalized angular momentum operator Λ^2 is defined as

$$\Lambda^2 := -\sum_{i>j}^{d} \sum_{j=1}^{d} \left(x_i \frac{\partial}{\partial x_j} - x_j \frac{\partial}{\partial x_i} \right)^2 \tag{10.3.19}$$

When $d = 3$ it reduces to the familiar orbital angular momentum operator

$$L^2 = L_1^2 + L_2^2 + L_3^2 \tag{10.3.20}$$

where

$$L_1 = \frac{1}{i} \left(x_2 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_2} \right) \tag{10.3.21}$$

and where L_2 and L_3 given by similar expressions with cyclic permutations of the coordinates.

10.3.5 Hyperspherical harmonics

If we expand the expression in (10.3.19), we obtain:

$$\Lambda^2 = -r^2\Delta + \sum_{i,j=1}^{d} x_i x_j \frac{\partial^2}{\partial x_i \partial x_j} + (d-1) \sum_{i=1}^{d} x_i \frac{\partial}{\partial x_i} \tag{10.3.22}$$

We next allow Λ^2 to act on a homogeneous polynomial f_n, and make use of (10.3.5). This give us:

$$\Lambda^2 f_n = -r^2 \Delta f_n + n(d-1) f_n + \sum_{i,j=1}^{d} x_i x_j \frac{\partial^2 f_n}{\partial x_i \partial x_j} \tag{10.3.23}$$

The relationship

$$\sum_{i,j=1}^{d} x_i x_j \frac{\partial^2 f_n}{\partial x_i \partial x_j} = n(n-1) f_n \tag{10.3.24}$$

can be derived in a manner similar to the derivation of (10.3.5). Substituting this into (10.3.24) we have:

$$\Lambda^2 f_n = -r^2 \Delta f_n + n(n+d-2) f_n \tag{10.3.25}$$

From (10.3.25) it follows that a harmonic polynomial of degree λ is an eigenfunction of the generalized angular momentum operator with the eigenvalue $\lambda(\lambda + d - 2)$, i.e.,

$$\Lambda^2 h_\lambda = \lambda(\lambda + d - 2) h_\lambda \tag{10.3.26}$$

When $d = 3$ this reduces to

$$L^2 h_l = l(l+1)h_l \tag{10.3.27}$$

We can conclude from this discussion that the canonical decomposition of a homogeneous polynomial can be viewed as a decomposition into eigenfunctions of generalized angular momentum.

Hyperspherical harmonics are defined by the relationship

$$Y_\lambda = \frac{1}{r^\lambda} h_\lambda \tag{10.3.28}$$

where r is the d-dimensional radius (10.3.7), and they are thus eigenfunctions of the generalized angular momentum operator Λ^2:

$$\Lambda^2 Y_\lambda = \lambda(\lambda + d - 2)Y_\lambda \tag{10.3.29}$$

For $d = 3$, this reduces to

$$\Lambda^2 Y_{lm} = l(l+1)Y_{lm} \tag{10.3.30}$$

The extra index m for the familiar three-dimensional spherical harmonics is due to an $(2l+1)$-fold degeneracy of the eigenfunctions for each eigenvalue. The three-dimensional spherical harmonics are usually classified according to the chain of groups $SO(3) \supset SO(2)$: The principal index l labels the eigenvalue of the angular momentum operator L^2, which is Λ^2 for $d = 3$, and the secondary index m labels the eigenvalue of L_z, which is Λ for $d = 2$. In the general case of higher dimensional spaces, we can classify the hyperspherical harmonics according to the chain of groups

$$SO(d) \supset SO(d-1) \subset \cdots SO(2) \tag{10.3.31}$$

with indices thus corresponding to the eigenvalues of

$$\Lambda_d^2, \Lambda_{d-1}^2, \ldots, \Lambda_3^2, \Lambda_2 \tag{10.3.32}$$

It can be shown, that in general the degeneracy of the hyperspherical harmonics belonging to the principal quantum number λ is given by

$$\frac{(d+2\lambda-3)(d+\lambda-3)!}{(\lambda-1)!(d-1)!} \tag{10.3.33}$$

10.3.6 Angular and hyperangular integration

In a 3-dimensional space the volume element is given by $dx_1 dx_2 dx_3$ in Cartesian coordinates or by $r^2 dr\, d\Omega$ in spherical polar coordinates. Thus we can write

$$dx_1 dx_2 dx_3 = r^2 dr\, d\Omega \tag{10.3.34}$$

where $d\Omega$ is the element of solid angle. Similarly, in a d-dimensional space we can write

$$dx_1 dx_2 \cdots dx_d = r^{d-1} dr\, d\Omega \tag{10.3.35}$$

where r is the hyperradius and where $d\Omega$ is the element of generalized solid angle. From the Hermiticity of the generalized angular momentum operator Λ^2, one can show that its eigenfunctions corresponding to different eigenvalues are orthogonal with respect to hyperangular integration. Thus from (10.3.26) it follows that

$$\int d\Omega\, h_{\lambda'}^* h_\lambda = 0 \quad \text{if } \lambda' \neq \lambda \tag{10.3.36}$$

In the particular case where $\lambda' = 0$, this becomes

$$\int d\Omega\, h_0^* h_\lambda = h_0^* \int d\Omega\, h_\lambda = 0 \quad \text{if } \lambda \neq 0 \tag{10.3.37}$$

since the constant, h_0^* can be factored out of the integration over generalized solid angle. Thus we obtain the important result:

$$\int d\Omega\, h_\lambda = 0 \quad \text{if } \lambda \neq 0 \tag{10.3.38}$$

Let us now combine this result with equation (10.3.11), which shows the form of the canonical decomposition of a homogeneous polynomial f_n. From (10.3.11) and (10.3.38) it follows that if a homogeneous polynomial is integrated over generalized solid angle, the only term that will survive is the constant term in the canonical decomposition, i.e., h_0. But this term, together with the power of the hyperradius multiplying it,can be factored out of the integration. Thus we obtain the powerful angular and hyperangular integration theorem:

$$\int d\Omega\, f_n = \begin{cases} r^n h_0 \int d\Omega & n = \text{even} \\ 0 & n = \text{odd} \end{cases} \tag{10.3.39}$$

where we have used the fact that when n is odd, the constant term h_0 does not occur in the canonical decomposition. We already have an explicit expression for h_0, namely equation (10.3.16). The only task remaining is to evaluate the total generalized solid angle, $\int d\Omega$. We can do this by comparing the integral of e^{-r^2} over the whole d-dimensional space when performed in Cartesian coordinates with the same integral performed in generalized spherical polar coordinates. Since the result must be the same, independent of the coordinate system used, we have:

$$\int_0^\infty dr\, r^{d-1} e^{-r^2} \int d\Omega = \prod_{j=1}^{d} \int_{-\infty}^{\infty} dx_j\, e^{-x_j^2} \tag{10.3.40}$$

The hyperradial integral can be expressed in terms of the gamma function:

$$\int_0^\infty dr\, r^{d-1} e^{-r^2} = \frac{\Gamma(d/2)}{2} \tag{10.3.41}$$

as can the integral over each of the Cartesian coordinates:

$$\int_{-\infty}^{\infty} dx_j\, e^{-x_j^2} = \Gamma(1/2) = \pi^{\frac{1}{2}} \tag{10.3.42}$$

Inserting these results into (10.3.40) and solving for $\int d\Omega$, we obtain:

$$\int d\Omega = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \tag{10.3.43}$$

Finally, combining (10.3.11), (10.3.39) and (10.3.43), we have an explicit expression for the integral over generalized solid angle of any homogeneous polynomial of degree n:

$$\int d\Omega\, f_n = \begin{cases} \dfrac{2\pi^{d/2} r^n (d-2)!!}{\Gamma(d/2) n!! (d+n-2)!!} \Delta^{\frac{1}{2}n} f_n & n = \text{even} \\ 0 & n = \text{odd} \end{cases} \tag{10.3.44}$$

Now suppose that $F(\mathbf{x})$ is any function whatever that can be expanded in a convergent series of homogeneous polynomials. If the series has the form:

$$F(\mathbf{x}) = \sum_{n=0}^{\infty} f_n(\mathbf{x}) \tag{10.3.45}$$

then it follows from (10.3.44) that

$$\int d\Omega\, F(\mathbf{x}) = \frac{(d-2)!!2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \sum_{n=0,2,..}^{\infty} \frac{r^n}{n!!(n+d-2)!!} \Delta^{n/2} f_n(\mathbf{x}) \tag{10.3.46}$$

We can notice that at the point $\mathbf{x} = 0$, all terms in a polynomial vanish, except the constant term. Thus we have

$$\left\lfloor \Delta^{n/2} F(\mathbf{x}) \right\rfloor_{\mathbf{x}=0} = \Delta^{n/2} f_n(\mathbf{x}) \tag{10.3.47}$$

This allows us to rewrite (10.3.46) in the form

$$\int d\Omega\, F(\mathbf{x}) = \frac{(d-2)!!2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \sum_{\nu=0}^{\infty} \frac{r^{2\nu}}{(2\nu)!!(d+2\nu-2)!!} \left\lfloor \Delta^{\nu} F(\mathbf{x}) \right\rfloor_{\mathbf{x}=0} \tag{10.3.48}$$

where we have made the substitution $n = 2\nu$. In the case where $d = 3$, this reduces to

$$\int d\Omega\, F(\mathbf{x}) = 4\pi \sum_{\nu=0}^{\infty} \frac{r^{2\nu}}{(2\nu+1)!} \left\lfloor \Delta^{\nu} F(\mathbf{x}) \right\rfloor_{\mathbf{x}=0} \tag{10.3.49}$$

10.3.7 An alternative method for angular and hyperangular integrations

The preceding method grows from the close relationship between harmonic polynomials and hyperspherical harmonics, and it works well when the degree of the polynomials is moderate. However, for polynomials of large degree, the memory consumption grows rapidly, and the method becomes impractical. The following method is efficient, even for polynomials of large degree. It is based on expressing the same d-dimensional integral both in Cartesian coordinates and in terms of an integral over the hyperradius multiplied by an integral over the generalized solid angle. The integral over Cartesian coordinates can be performed exactly, as can the integral over the hyperradius, and by this means, the hyperangular integral is determined.

Theorem 10.3.1.

Let

$$I(\mathbf{n}) := \int d\Omega \left(\frac{x_1}{r}\right)^{n_1} \left(\frac{x_2}{r}\right)^{n_2} ..... \left(\frac{x_d}{r}\right)^{n_d} \tag{10.3.50}$$

where $x_1, x_2, ..., x_d$ are the Cartesian coordinates of a d-dimensional space, $d\Omega$ is the generalized solid angle, r is the hyperradius, defined by

$$r^2 := \sum_{j=1}^{d} x_j^2 \tag{10.3.51}$$

and where the n_j's are positive integers or zero. Then

$$I(\mathbf{n}) = \begin{cases} \dfrac{\pi^{d/2}}{2^{(n/2-1)}\Gamma\left(\frac{d+n}{2}\right)} \displaystyle\prod_{j=1}^{d} (n_j - 1)!! & \text{if all the } n_j\text{'s are even} \\ 0 \qquad \text{otherwise} \end{cases} \tag{10.3.52}$$

where

$$n := \sum_{j=1}^{d} n_j \tag{10.3.53}$$

Proof. To prove this theorem, we consider the integral

$$\int_0^\infty dr\, r^{d-1} e^{-r^2} \int d\Omega\, x_1^{n_1} x_2^{n_2} \ldots\ldots\ldots x_d^{n_d} = \prod_{j=1}^{d} \int_{-\infty}^{\infty} dx_j\, x_j^{n_j} e^{-x_j^2} \tag{10.3.54}$$

If n_j is zero or a positive integer, then

$$\int_{-\infty}^{\infty} dx_j\, x_j^{n_j} e^{-x_j^2} = \begin{cases} \dfrac{(n_j-1)!!\sqrt{\pi}}{2^{n_j/2}} & \text{if } n_j \text{ is even} \\ 0 \qquad \text{if } n_j \text{ is odd} \end{cases} \tag{10.3.55}$$

so that the right-hand side of 10.3.54 becomes

$$\prod_{j=1}^{d} \int_{-\infty}^{\infty} dx_j\, x_j^{n_j} e^{-x_j^2} = \begin{cases} \dfrac{\pi^{d/2}}{2^{n/2}} \displaystyle\prod_{j=1}^{d} (n_j-1)!! & \text{if all the } n_j\text{'s are even} \\ 0 \qquad \text{otherwise} \end{cases} \tag{10.3.56}$$

The left-hand side of (10.3.54) can be written in the form

$$\int_0^\infty dr\, r^{d+n-1} e^{-r^2} \int d\Omega \left(\frac{x_1}{r}\right)^{n_1} \left(\frac{x_2}{r}\right)^{n_2} \ldots\ldots\ldots \left(\frac{x_d}{r}\right)^{n_d} = \frac{I(\mathbf{n})}{2} \Gamma\left(\frac{d+n}{2}\right) \tag{10.3.57}$$

Substituting (10.3.56) and (10.3.57) into (10.3.54), we obtain (10.3.52), QED. □

In the special case where $d = 3$, equation (10.3.52) becomes

$$\int d\Omega \left(\frac{x_1}{r}\right)^{n_1} \left(\frac{x_2}{r}\right)^{n_2} \left(\frac{x_d}{r}\right)^{n_3} = \begin{cases} \dfrac{4\pi}{(n+1)!!} \displaystyle\prod_{j=1}^{3} (n_j-1)!! & \text{if all the } n_j\text{'s are even} \\ 0 \qquad \text{otherwise} \end{cases} \tag{10.3.58}$$

Let us now consider a general polynomial (not necessarily homogeneous) of the form:

$$P(\mathbf{x}) = \sum_{\mathbf{n}} c_{\mathbf{n}}\, x_1^{n_1} x_2^{n_2} \ldots\ldots\ldots x_d^{n_d} \tag{10.3.59}$$

Then we have

$$\int d\Omega\, P(\mathbf{x}) = \sum_{\mathbf{n}} c_{\mathbf{n}} \int d\Omega\, x_1^{n_1} x_2^{n_2} \ldots\ldots\ldots x_d^{n_d} = \sum_{\mathbf{n}} c_{\mathbf{n}}\, r^n I(\mathbf{n}) \tag{10.3.60}$$

It can be seen that equation (10.3.52) can be used to evaluate the generalized angular integral of any polynomial whatever, regardless of whether or not it is homogeneous.

It is interesting to ask what happens if the n_j's are not required to be zero or positive integers. If all the n_j's are real numbers greater than -1, then the right-hand side of (10.3.54) can still be evaluated and it has the form

$$\prod_{j=1}^{d} \int_{-\infty}^{\infty} dx_j\, x_j^{n_j} e^{-x_j^2} = \prod_{j=1}^{d} \frac{1}{2} \left(1 + e^{i\pi n_j}\right) \Gamma\left(\frac{n_j+1}{2}\right) \tag{10.3.61}$$

Thus (10.3.52) becomes

$$I(\mathbf{n}) = \frac{2}{\Gamma\left(\frac{d+n}{2}\right)} \prod_{j=1}^{d} \frac{1}{2} \left(1 + e^{i\pi n_j}\right) \Gamma\left(\frac{n_j+1}{2}\right) \qquad n_j > -1, \qquad j = 1, \ldots, d \tag{10.3.62}$$

This more general equation reduces to 10.3.52 in the special case where the n_j's are required to be either zero or positive integers.

10.3.8 Angular integrations by a vector-pairing method

The third method presented depends on the invariance under rotation of scalar products. An integral, which is just a number, must be independent of the coordinate system, and therefore it be expressed in quantities that are invariant under rotation of the coordinate system. In $\mathbb{R}^d$, this means that they must be expressed as scalar products.

Let us consider the following integral in a 3-dimensional space:

$$I = \frac{1}{4\pi} \int d\Omega \, (\hat{\mathbf{x}} \cdot \hat{\mathbf{A}})(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) \tag{10.3.63}$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are unit vectors. Since the integral I must be independent of $\mathbf{x}$ and invariant under rotations, it must be proportional to the scalar product, $\hat{\mathbf{A}} \cdot \hat{\mathbf{B}}$, which is the only scalar that can be made out of two vectors. The constant of proportionality can be found by considering the case where $\hat{\mathbf{A}} = \hat{\mathbf{B}}$, and in this way, one finds that

$$\frac{1}{4\pi} \int d\Omega \, (\hat{\mathbf{x}} \cdot \hat{\mathbf{A}})(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) = \frac{1}{3}(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}}) \tag{10.3.64}$$

Building on this approach to angular integration, Avery, Ørmen and Michels (Avery and Ørmen [1980], Michels [1981]) were able to show that

$$\begin{aligned} &\frac{1}{4\pi} \int d\Omega \, (\hat{\mathbf{x}} \cdot \hat{\mathbf{A}})^{n_1} (\hat{\mathbf{x}} \cdot \hat{\mathbf{B}})^{n_2} \\ &= \frac{n_1! n_2!}{(n_1 + n_2 + 1)!!} \sum_{\lambda=1,3,5\ldots}^{m} \frac{(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^{\lambda}}{\lambda!(n_1 - \lambda)!!(n_2 - \lambda)!!} \end{aligned} \tag{10.3.65}$$

where m is equal either to n_1 or to n_2, whichever is the smaller. For the case involving three unit vectors, $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$, one obtains

$$\begin{aligned} &\frac{1}{4\pi} \int d\Omega \, (\hat{\mathbf{x}} \cdot \hat{\mathbf{A}})^{n_1} (\hat{\mathbf{x}} \cdot \hat{\mathbf{B}})^{n_2} (\hat{\mathbf{x}} \cdot \hat{\mathbf{C}})^{n_3} \\ &= \frac{n_1! n_2! n_3!}{(n_1 + n_2 + + n_3 + 1)!!} \sum_{\lambda_{12}\lambda_{23}\lambda_{13}} \\ &\times \frac{(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^{\lambda_{12}} (\hat{\mathbf{A}} \cdot \hat{\mathbf{C}})^{\lambda_{13}} (\hat{\mathbf{B}} \cdot \hat{\mathbf{C}})^{\lambda_{23}}}{\lambda_{12}!\lambda_{23}!\lambda_{13}!(n_1 - \lambda_{12} - \lambda_{13})!!(n_2 - \lambda_{12} - \lambda_{23})!!(n_3 - \lambda_{13} - \lambda_{23})!!} \end{aligned} \tag{10.3.66}$$

where each λ_{ij}'s runs from zero to either n_i or n_j, whichever the smaller, and where they must also satisfy the following criteria:

$$\begin{aligned} \lambda_{12} + \lambda_{13} &\leq n_1 \qquad & (-1)^{\lambda_{12}+\lambda_{13}} &= (-1)^{n_1} \\ \lambda_{12} + \lambda_{23} &\leq n_2 \qquad & (-1)^{\lambda_{12}+\lambda_{23}} &= (-1)^{n_2} \\ \lambda_{13} + \lambda_{23} &\leq n_3 \qquad & (-1)^{\lambda_{13}+\lambda_{23}} &= (-1)^{n_3} \end{aligned} \tag{10.3.67}$$

If these criteria are not fulfilled by the λ_{ij}'s, the term is not included in the sum. More generally,

$$\begin{aligned} &\frac{1}{4\pi} \int d\Omega \, \prod_{j=1}^{N} (\hat{\mathbf{x}} \cdot \hat{\mathbf{A}}_j)^{n_j} \\ &= \frac{1}{(n_1 + n_2 + \ldots + n_N + 1)!!} \sum_{\lambda^*} \prod_{j=1}^{N} \frac{n_j!}{D_j} \prod_{i=j}^{N} (\hat{\mathbf{A}}_i \cdot \hat{\mathbf{A}}_j)^{\lambda_{ij}} \end{aligned} \tag{10.3.68}$$

where

$$D_j = \left(n_j - \sum_{i=1}^{j-1} \lambda_{ij} - \sum_{i=j+1}^{N} \lambda_{ij} \right)!! \prod_{i=j+1}^{N} \lambda_{ij}! \tag{10.3.69}$$

In (10.3.68) the sum $\sum_{\lambda^*}$ denotes a sum over all values of λ_{ij} which are positive integers or zero and which fulfil the criteria

$$\sum_{i=1}^{j} \lambda_{ij} + \sum_{i=j}^{N} \lambda_{ji} = n_j \qquad\qquad j = 1, 2, ..., N \tag{10.3.70}$$

It is easy to extend these methods to spaces of higher dimension, and the relevant formulae can be found in references Avery and Ørmen [1980] and Michels [1981]. It is also possible to evaluate integrals of the type

$$W_{l,l',l''} := \frac{1}{4\pi} \int d\Omega \; P_l(\hat{\mathbf{x}} \cdot \hat{\mathbf{A}}) P_{l'}(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) P_{l''}(\hat{\mathbf{x}} \cdot \hat{\mathbf{C}}) \tag{10.3.71}$$

and some examples are shown in the following table, where only non-zero values are shown. In order for $W_{l,l',l''}$ to be nonzero, $l + l' + l''$ must be even and $|l - l'| \leq l'' \leq l + l'$.

(l, l', l'')	$W_{l,l',l''} := \frac{1}{4\pi} \int d\Omega \; P_l(\hat{\mathbf{x}} \cdot \hat{\mathbf{A}}) P_{l'}(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) P_{l''}(\hat{\mathbf{x}} \cdot \hat{\mathbf{C}})$
$(0,0,0)$	1
$(1,1,0)$	$\frac{1}{3}(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})$
$(1,1,2)$	$\frac{1}{15}\left[-(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}}) + 3(\hat{\mathbf{A}} \cdot \hat{\mathbf{C}})(\hat{\mathbf{B}} \cdot \hat{\mathbf{C}})\right]$
$(2,2,0)$	$\frac{1}{10}\left[-1 + 3(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^2\right]$
$(2,2,2)$	$\frac{1}{35}\left[2 - 3(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^2 - 3(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^2 - 3(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^2 + 9(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})(\hat{\mathbf{A}} \cdot \hat{\mathbf{C}})(\hat{\mathbf{B}} \cdot \hat{\mathbf{C}})\right]$
$(1,2,3)$	$\frac{3}{70}\left[-(\hat{\mathbf{A}} \cdot \hat{\mathbf{C}}) - 2(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})(\hat{\mathbf{B}} \cdot \hat{\mathbf{C}}) + 5(\hat{\mathbf{A}} \cdot \hat{\mathbf{C}})(\hat{\mathbf{B}} \cdot \hat{\mathbf{C}})^2\right]$
$(3,3,0)$	$\frac{1}{14}\left[-3(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}}) + 5(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^3\right]$

10.4 Hyperspherical harmonics and Fourier transformed Coulomb Sturmians

10.4.1 Many-center Coulomb Sturmians

Shibuya-Wulfman integrals occur when Coulomb Sturmian basis sets are used to construct molecular orbitals. The one-electron equation obeyed by a molecular orbital $\varphi(\mathbf{x})$ has the form

$$\left[-\frac{1}{2}\nabla^2 + v(\mathbf{x}) - \epsilon_j\right] \varphi(\mathbf{x}) = 0 \tag{10.4.1}$$

If the potential $v(\mathbf{x})$ is taken to be the nuclear attraction potential of the bare nuclei, then

$$v(\mathbf{x}) = -\sum_a \frac{Z_a}{|\mathbf{x} - \mathbf{x}_a|} \tag{10.4.2}$$

We can try to represent the molecular orbital as a superposition of many-center Coulomb Sturmians centered on the various nuclei of the molecule

$$\varphi(\mathbf{x}) = \sum_\tau \chi_\tau(\mathbf{x}) C_\tau \tag{10.4.3}$$

Here

$$\chi_\tau(\mathbf{x}) := \chi_{nlm}(\mathbf{x} - \mathbf{x}_a) \tag{10.4.4}$$

represents a Coulomb Sturmian of the type discussed in Section 10.1, centered on a nucleus at the point $\mathbf{x}_a$, while

$$\tau := (n, l, m, a) \tag{10.4.5}$$

is a set of four indices, the last index being that of the nucleus on which the Coulomb Sturmian is centered. The Sturmian basis functions are isoenergetic, and we let the common energy of all the members of the basis set correspond to that of the molecular orbital that we wish to represent:

$$\epsilon := -\frac{1}{2}k^2 \tag{10.4.6}$$

Substituting these relationships into equation (10.4.1) and taking the scalar product with a conjugate member of the basis set, we obtain the secular equations:

$$\sum_\tau \int dx \chi_{\tau'}(\mathbf{x}) \left[-\frac{1}{2}\nabla^2 + \frac{1}{2}k^2 + v(\mathbf{x})\right] \chi_\tau(\mathbf{x}) C_\tau = 0 \tag{10.4.7}$$

10.4.2 Definition of Shibuya-Wulfman integrals

We now define the Shibuya-Wulfman integrals $\mathfrak{S}_{\tau',\tau}$ by the relationship

$$\mathfrak{S}_{\tau',\tau} := \frac{1}{k^2} \int d^3x \, \chi_{\tau'}(\mathbf{x}) \left[-\frac{1}{2}\nabla^2 + \frac{1}{2}k^2\right] \chi_\tau(\mathbf{x}) \tag{10.4.8}$$

while the Wulfman integrals $\mathfrak{W}_{\tau',\tau}$ (integrals of the potential) are defined by

$$\mathfrak{W}_{\tau',\tau} := -\frac{1}{k} \int d^3x \, \chi_{\tau'}(\mathbf{x}) v(\mathbf{x}) \chi_\tau(\mathbf{x}) \tag{10.4.9}$$

With this notation, the secular equations can be written as

$$\sum_\tau \left[\mathfrak{W}_{\tau',\tau} - k\mathfrak{S}_{\tau',\tau}\right] C_\tau = 0 \tag{10.4.10}$$

Notice that the roots of the secular equations are not energies but values of the scaling parameter k, which are related to the orbital energies through equation (10.4.6).

The Shibuya-Wulfman integrals were first introduced by T. Shibuya and C. Wulfman in 1965 in connection with their famous momentum-space treatment of many-center one-electron problems (Shibuya and Wulfman [1965]). These integrals can conveniently be evaluated in momentum space using the Fock projection, which establishes a relationship between hyperspherical harmonics and the Fourier transforms of Coulomb Sturmians. The problem of evaluating these integrals, as well as many other integrals needed in molecular problems, can then be converted into a problem of hyperangular integration, as is shown in the next section.

10.5 Shibuya-Wulfman integrals and Sturmian overlap integrals evaluated in terms of hyperspherical harmonics

The Shibuya-Wulfman integrals $\mathfrak{S}_{\tau',\tau}$ defined by equation (14.1.13), as well as the molecular Sturmian overlap integrals

$$m_{\tau',\tau} := \int d^3x\, \chi^*_{\tau'}(\mathbf{x})\chi_\tau(\mathbf{x}) \tag{10.5.1}$$

can conveniently be evaluated in reciprocal space. Using the fact that

$$m_{\tau',\tau} := \int d^3x\, \chi^*_{\tau'}(\mathbf{x})\chi_\tau(\mathbf{x}) = \int d^3p\, \chi^{t*}_{\tau'}(\mathbf{p})\chi^t_\tau(\mathbf{p}) \tag{10.5.2}$$

where

$$\begin{aligned} \chi^t_\tau(\mathbf{p}) &:= \frac{1}{(2\pi)^{3/2}} \int d^3x\, e^{-i\mathbf{p}\cdot\mathbf{x}}\chi_\tau(\mathbf{x}) = e^{-i\mathbf{p}\cdot\mathbf{x}_a}\chi^t_\mu(\mathbf{p}) \\ \chi^{t*}_{\tau'}(\mathbf{p}) &:= \frac{1}{(2\pi)^{3/2}} \int d^3x\, e^{i\mathbf{p}\cdot\mathbf{x}}\chi^*_{\tau'}(\mathbf{x}) = e^{i\mathbf{p}\cdot\mathbf{x}_{a'}}\chi^{t*}_\mu(\mathbf{p}) \end{aligned} \tag{10.5.3}$$

we obtain

$$m_{\tau',\tau} = \int d^3p\, e^{i\mathbf{p}\cdot\mathbf{R}}\chi^{t*}_{\mu'}(\mathbf{p})\chi^t_\mu(\mathbf{p}) \tag{10.5.4}$$

where

$$\mathbf{R} := \mathbf{x}_{a'} - \mathbf{x}_a \tag{10.5.5}$$

We now make use of V. Fock's relationship

$$\chi^t_\mu(\mathbf{p}) = M(p)Y_{n-1.l.m}(\mathbf{u}) \tag{10.5.6}$$

Here

$$M(p) := \frac{4k^{5/2}}{(k^2+p^2)^2} \tag{10.5.7}$$

while $Y_\mu(\mathbf{u})$ is a 4-dimensional hyperspherical harmonic defined by:

$$Y_{\lambda,l,m}(\mathbf{u}) = \mathcal{N}_{\lambda,l}C^{1+l}_{\lambda-l}(u_4)Y_{l,m}(u_1,u_2,u_3) \tag{10.5.8}$$

where

$$\mathcal{N}_{\lambda,l} = (-1)^\lambda i^l (2l)!! \sqrt{\frac{2(\lambda+1)(\lambda-l)!}{\pi(\lambda+l+1)!}} \tag{10.5.9}$$

is a normalizing constant while

$$C_n^a(u_4) = \sum_{t=0}^{[n/2]} \frac{(-1)^t \Gamma(n+a-t)}{t!(n-2t)!\Gamma(a)} (2u_4)^{n-2t} \tag{10.5.10}$$

is a Gegenbauer polynomial, and where $Y_{l,m}$ is a familiar 3-dimensional spherical harmonic. In equation (10.5.6), $\mathbf{u}$ is a 4-dimensional unit vector that defines Fock's projection of momentum space onto the surface of a 4-dimensional hypersphere.

$$\mathbf{u} = (u_1, u_2, u_3, u_4) = \left(\frac{2kp_1}{k^2+p^2}, \frac{2kp_2}{k^2+p^2}, \frac{2kp_3}{k^2+p^2}, \frac{k^2-p^2}{k^2+p^2} \right) \tag{10.5.11}$$

Substituting (10.5.6) into (10.5.4), we obtain

$$\begin{aligned} m_{\tau',\tau} &= \int d^3p \, e^{i\mathbf{p}\cdot\mathbf{R}} M(p)^2 Y^*_{n'-1,l',m'}(\mathbf{u}) Y_{n-1,l,m}(\mathbf{u}) \\ &:= \int d^3p \, e^{i\mathbf{p}\cdot\mathbf{R}} M(p)^2 Y^*_{\mu'}(\mathbf{u}) Y_{\mu}(\mathbf{u}) \end{aligned} \tag{10.5.12}$$

One can show that the Shibuya-Wulfman integrals can be written in a similar form:

$$\mathfrak{S}_{\tau',\tau} = \int d^3p \, e^{i\mathbf{p}\cdot\mathbf{R}} \left(\frac{2k}{k^2+p^2} \right)^3 Y^*_{\mu'}(\mathbf{u}) Y_{\mu}(\mathbf{u}) \tag{10.5.13}$$

One can also show that

$$\int d^3p \, e^{i\mathbf{p}\cdot\mathbf{R}} \left(\frac{2k}{k^2+p^2} \right)^3 Y^*_{\mu}(\mathbf{u}) = (2\pi)^{3/2} f_{n,l}(s) Y_{l,m}(\hat{\mathbf{s}}) \tag{10.5.14}$$

where $Y_{l,m}$ is an ordinary 3-dimensional spherical harmonic and where

$$\mathbf{s} = \{s_x, s_y, s_z\} := k\mathbf{R} \qquad s = k|\mathbf{R}| \tag{10.5.15}$$

The function $f_{n,l}(s)$ is defined by

$$\begin{aligned} k^{3/2} f_{n,l}(s) &:= R_{n,l}(s) - \frac{1}{2}\sqrt{\frac{(n-l)(n+l+1)}{n(n+1)}} R_{n+1,l}(s) \\ &\quad - \frac{1}{2}\sqrt{\frac{(n+l)(n-l-1)}{n(n-1)}} R_{n-1,l}(s) \end{aligned} \tag{10.5.16}$$

where $R_{n,l}$ is the radial function of the Coulomb Sturmians given in Equation (10.1.2), and where

$$R_{n-1,l}(s) := 0 \qquad \text{if } l > n-1 \tag{10.5.17}$$

Similarly, one can show (Avery and Avery [2006]) that

$$\int d^3p \, e^{i\mathbf{p}\cdot\mathbf{R}} M(p)^2 Y_{\mu}(\mathbf{u}) = (2\pi)^{3/2} g_{n,l}(s) Y_{l,m}(\hat{\mathbf{s}}) \tag{10.5.18}$$

where

$$\begin{aligned} g_{n,l}(s) &:= f_{n,l}(s) - \frac{1}{2}\sqrt{\frac{(n-l)(n+l+1)}{n(n+1)}} f_{n+1,l}(s) \\ &\quad - \frac{1}{2}\sqrt{\frac{(n+l)(n-l-1)}{n(n-1)}} f_{n-1,l}(s) \end{aligned} \tag{10.5.19}$$

where we define

$$f_{n-1,l}(s) := 0 \qquad \text{if } l > n-1 \tag{10.5.20}$$

Equations (15.1.7) and (10.5.18) are respectively identical with the Shibuya Wulfman integrals and the molecular Sturmian overlap integrals except that they contain only one 4-dimensional hyperspherical harmonic instead of a product of two. Thus the problem of evaluating both $\mathfrak{S}_{\tau',\tau}$ and $m_{\tau',\tau}$ reduces to the problem of evaluating the coefficients

$$c_{\mu'';\mu',\mu} = \int d\Omega_4 Y^*_{\mu''}(\mathbf{u}) Y^*_{\mu'}(\mathbf{u}) Y_{\mu}(\mathbf{u}) \tag{10.5.21}$$

These coefficients can readily be pre-evaluated once and for all using the hyperangular integration theorems discussed in Section 10.3, and they can be stored as a large but very sparse matrix. We then obtain the relationships:

$$Y^*_{\mu'}(\mathbf{u}) Y_{\mu}(\mathbf{u}) = \sum_{\mu''} Y_{\mu''}(\mathbf{u}) c_{\mu'';\mu',\mu} \tag{10.5.22}$$

$$\mathfrak{S}_{\tau',\tau} = (2\pi)^{3/2} \sum_{\mu''} Y_{l'',m''}(\hat{\mathbf{s}}) f_{n'',l''}(s) c_{\mu'';\mu',\mu} \tag{10.5.23}$$

and

$$m_{\tau',\tau} = (2\pi)^{3/2} \sum_{\mu''} Y_{l'',m''}(\hat{\mathbf{s}}) g_{n'',l''}(s) c_{\mu'';\mu',\mu} \tag{10.5.24}$$

10.6 Asymptotic behaviour: The multipole expansion

The electrostatic interaction between a charge density ρ_a, centered at the point $\mathbf{x}_a$, and another charge density ρ_b, centered at the point $\mathbf{x}_b$, is given by

$$J = \int d^3x \int d^3x' \, \rho_a(\mathbf{x} - \mathbf{x}_a) \rho_b(\mathbf{x}' - \mathbf{x}_b) \frac{1}{|\mathbf{x} - \mathbf{x}'|} \tag{10.6.1}$$

If we introduce the notation

$$\begin{aligned} \mathbf{x}_a &:= \mathbf{x} - \mathbf{x}_a \\ \mathbf{x}_b &:= \mathbf{x}' - \mathbf{x}_b \\ \mathbf{R} &:= \mathbf{x}_b - \mathbf{x}_a \\ R &:= |\mathbf{x}_b - \mathbf{x}_a| \\ \mathbf{x}_{ab} &:= \mathbf{x}_a - \mathbf{x}_b \\ r_{ab} &:= |\mathbf{x}_a - \mathbf{x}_b| \end{aligned} \tag{10.6.2}$$

we can rewrite the interelectron repulsion integral as

$$\begin{aligned} J &= \int d^3x \int d^3x' \, \rho_a(\mathbf{x} - \mathbf{x}_a) \rho_b(\mathbf{x}' - \mathbf{x}_b) \frac{1}{|\mathbf{x} - \mathbf{x}'|} \\ &= \int d^3x_a \int d^3x_b \, \rho_a(\mathbf{x}_a) \rho_b(\mathbf{x}_b) \frac{1}{|\mathbf{x}_a - \mathbf{x}_b + \mathbf{x}_a - \mathbf{x}_b|} \\ &= \int d^3x_a \int d^3x_b \, \rho_a(\mathbf{x}_a) \rho_b(\mathbf{x}_b) \frac{1}{|\mathbf{x}_{ab} - \mathbf{R}|} \\ &= \int d^3x_a \int d^3x_b \, \rho_a(\mathbf{x}_a) \rho_b(\mathbf{x}_b) \sum_{l=0}^{\infty} \frac{r_<^l}{r_>^{l+1}} P_l(\hat{\mathbf{R}} \cdot \hat{\mathbf{x}}_{ab}) \end{aligned} \tag{10.6.3}$$

where

$$\frac{r_<^l}{r_>^{l+1}} := \begin{cases} r_{ab}^l / R^{l+1} & r_{ab} < R \\ R^l / r_{ab}^{l+1} & r_{ab} > R \end{cases} \tag{10.6.4}$$

In the asymptotic region, where the charge distributions do not overlap appreciably, we can assume that $r_{ab} < R$, and therefore we can write

$$\begin{aligned} J &\rightarrow \sum_{l=0}^{\infty} \frac{1}{R^{l+1}} \int d^3x_a \int d^3x_b\, \rho_a(\mathbf{x}_a)\rho_b(\mathbf{x}_b) r_{ab}^l P_l(\hat{\mathbf{R}} \cdot \hat{\mathbf{x}}_{ab}) \\ &= \sum_{l=0}^{\infty} \frac{1}{R^{l+1}} \int d^3x_a \int d^3x_b\, \rho_a(\mathbf{x}_a)\rho_b(\mathbf{x}_b) r_{ab}^l \sum_{m=-l}^{l} \frac{4\pi}{2l+1} Y_{l,m}^*(\hat{\mathbf{R}}) Y_{l,m}(\hat{\mathbf{x}}_{ab}) \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{R^{l+1}} \frac{4\pi}{2l+1} Y_{l,m}^*(\hat{\mathbf{R}}) \int d^3x_a \int d^3x_b\, \rho_a(\mathbf{x}_a)\rho_b(\mathbf{x}_b) r_{ab}^l Y_{l,m}(\hat{\mathbf{x}}_{ab}) \end{aligned} \tag{10.6.5}$$

We now introduce the regular and irregular solid harmonics R_l^m and I_l^m, defined by

$$\begin{aligned} \mathsf{R}_l^m(\mathbf{x}) &:= \sqrt{\frac{4\pi}{2l+1}} r^l Y_{lm}(\hat{\mathbf{x}}) \\ \mathsf{I}_l^m(\mathbf{x}) &:= \sqrt{\frac{4\pi}{2l+1}} \frac{1}{r^{l+1}} Y_{lm}(\hat{\mathbf{x}}) \end{aligned} \tag{10.6.6}$$

In terms of these, the asymptotic expression for J becomes

$$J \rightarrow \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^m \mathsf{I}_l^{-m}(\mathbf{R}) \int d^3x_a \int d^3x_b\, \rho_a(\mathbf{x}_a)\rho_b(\mathbf{x}_b) \mathsf{R}_l^m(\mathbf{x}_{ab}) \tag{10.6.7}$$

The double integral $\int d^3x_a \int d^3x_b$ can be separated into a product of single integrals by means of a standard expansion that makes use of binomial coefficients and Clebsch-Gordan coefficients:

$$\begin{aligned} &\mathsf{R}_l^m(\mathbf{x}_a - \mathbf{x}_b) \\ &= \sum_{l_a=0}^{l} (-1)^{l-l_a} \begin{pmatrix} 2l \\ 2l_a \end{pmatrix}^{1/2} \sum_{m_a=-l_a}^{l_a} \mathsf{R}_{l_a}^{m_a}(\mathbf{x}_a) \mathsf{R}_{l-l_a}^{m-m_a}(\mathbf{x}_b) \begin{pmatrix} l_a & l-l_a & l \\ m_a & m-m_a & m \end{pmatrix} \end{aligned} \tag{10.6.8}$$

Finally, making the substitutions $\rho_a(\mathbf{x}_a) = \rho_{\mu_1,\mu_2}(\mathbf{x}_a)$ and $\rho_a(\mathbf{x}_b) = \rho_{\mu_3,\mu_4}(\mathbf{x}_b)$, we obtain the asymptotic expression

$$\begin{aligned} &J_{\mu_1,\mu_2;\mu_3,\mu_4} \\ &\rightarrow \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^m \mathsf{I}_l^{-m}(\mathbf{R}) \sum_{l_a=0}^{l} (-1)^{l-l_a} \begin{pmatrix} 2l \\ 2l_a \end{pmatrix}^{1/2} \begin{pmatrix} l_a & l-l_a & l \\ m_a & m-m_a & m \end{pmatrix} \\ &\times \int d^3x_a\, \rho_{\mu 1,\mu 2}(\mathbf{x}_a) \mathsf{R}_{l_a}^{m_a}(\mathbf{x}_a) \int d^3x_b\, \rho_{\mu_3,\mu_4}(\mathbf{x}_b) \mathsf{R}_{l-l_a}^{m-m_a}(\mathbf{x}_b) \end{aligned} \tag{10.6.9}$$

In the special case where the densities $\rho_a(\mathbf{x}_a) = \rho_{\mu_1,\mu_2}(\mathbf{x}_a)$ and $\rho_a(\mathbf{x}_b) = \rho_{\mu_3,\mu_4}(\mathbf{x}_b)$ are formed by products of Coulomb Sturmian basis functions, this can be rewritten in the form:

$$\begin{aligned} &J_{\mu_1,\mu_2;\mu_3,\mu_4} \\ &\rightarrow k \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^m \mathsf{I}_l^{-m}(\mathbf{S}) \sum_{l_a=0}^{l} (-1)^{l-l_a} \begin{pmatrix} 2l \\ 2l_a \end{pmatrix}^{1/2} \begin{pmatrix} l_a & l-l_a & l \\ m_a & m-m_a & m \end{pmatrix} \\ &\times \int d^3s_a\, \rho_{\mu 1,\mu 2}(\mathbf{s}_a) \mathsf{R}_{l_a}^{m_a}(\mathbf{s}_a) \int d^3s_b\, \rho_{\mu_3,\mu_4}(\mathbf{x}_b) \mathsf{R}_{l-l_a}^{m-m_a}(\mathbf{s}_b) \end{aligned} \tag{10.6.10}$$

where $\mathbf{s}_a := k\mathbf{x}_a$, $\mathbf{s}_b := k\mathbf{x}_b$ and $\mathbf{S} := k\mathbf{R}$.

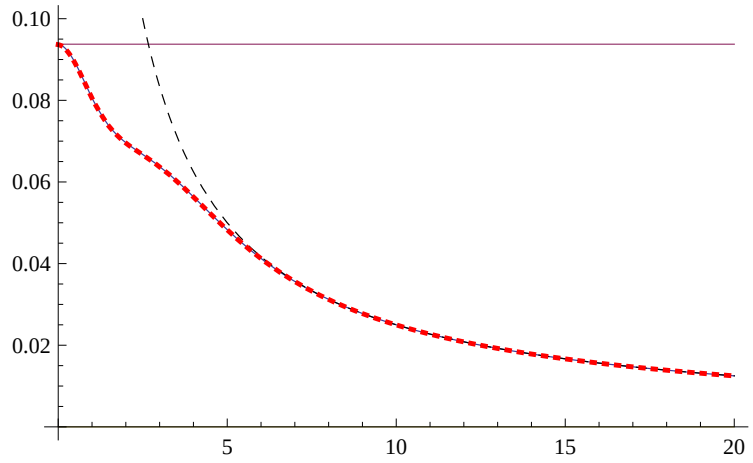


Figure 10.1: In this figure, $J_{1s,2s,2s,1s}$ and its asymptotic curve is shown as a function of $S = kR$ for a diatomic molecule, where R is the internuclear distance. The asymptotic curve is shown as a thin dashed line. For comparison, the atomic value of $J_{1s,2s,2s,1s}$ is shown as a thin solid line, while the integral calculated using the hyperspherical method of Chapter 11 is shown as a line with thick dashes and approximated using the method of Chapter 12 shown as a thin blue line. We see that there is good agreement between the methods in this case.

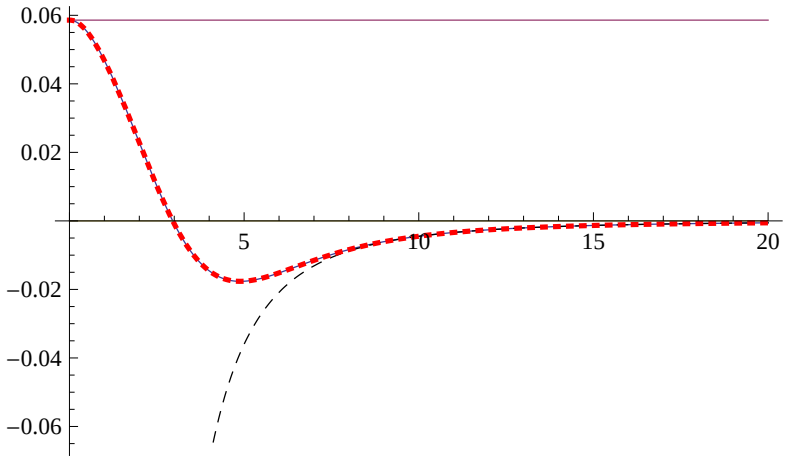


Figure 10.2: This figure is the same as Figure 10.1 except that $J_{2p_0,2s,2p_0,2s}$ is shown.

Integral	Asymptote	Integral	Asymptote	Integral	Asymptote
$J_{1s,1s,1s,1s}$	$\frac{1}{S}$	$J_{2p_0,2p_0,2p_0,2p_0}$	$\frac{6}{S^3}+\frac{1}{S}$	$J_{2p_{\pm1},2p_{\pm1},2p_{\pm1},2p_{\pm1}}$	$-\frac{3}{S^3}+\frac{1}{S}$
$J_{1s,1s,2s,2s}$	$\frac{1}{S}$	$J_{2s,2s,2p_0,2p_0}$	$\frac{3}{S^3}+\frac{1}{S}$	$J_{2s,2s,2p_{\pm1},2p_{\pm1}}$	$-\frac{3}{2S^3}+\frac{1}{S}$
$J_{2s,2s,2s,2s}$	$\frac{1}{S}$	$J_{1s,1s,1s,2p_0}$	$-\frac{1}{S^2}$	$J_{1s,1s,1s,2p_{\pm1}}$	$-\frac{1}{S^2}$
$J_{1s,1s,1s,2s}$	$-\frac{1}{2S}$	$J_{1s,2p_0,1s,2p_0}$	$-\frac{2}{S^3}$	$J_{1s,2p_{\pm1},1s,2p_{\pm1}}$	0
$J_{1s,2s,2s,2s}$	$-\frac{1}{2S}$	$J_{2s,2s,2p_0,2p_0}$	$\frac{3}{S^3}+\frac{1}{S}$	$J_{2s,2s,2p_{\pm1},2p_{\pm1}}$	$-\frac{3}{2S^3}+\frac{1}{S}$
$J_{1s,2s,1s,2s}$	$\frac{1}{4S}$	$J_{2s,2p_0,2s,2p_0}$	$-\frac{9}{2S^2}$	$J_{2s,2p_{\pm1},2s,2p_{\pm1}}$	0

Table 10.1: For diatomic systems, the interelectron repulsion integrals are pure functions of the scaled internuclear separation distance S, and the multipole expansions become particularly simple. This table lists the asymptotic expressions for a few integrals with the constellation $(a_1, a_2, a_3, a_4) = (0, 0, \mathbf{S}, \mathbf{S})$. Note that for diatomics, this is the only interesting type of integrals, since for all mixed-density integrals, the asymptotic expression is identically zero as the functions vanish at long ranges.

10.7 Expansion of displaced functions in terms of Legendre polynomials

10.7.1 Displaced spherically symmetric potentials

When a potential is a pure function of the distance between two particles, it is always possible to make an expansion of the form:

$$V(|\mathbf{x}_1 - \mathbf{x}_2|) = \sum_{l=0}^{\infty} v_l(r_1, r_2) P_l(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) \tag{10.7.1}$$

We would like to calculate the functions $v_l(r_1, r_2)$. To do this, we multiply (10.7.1) from the left by $P_{l'}(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2)$ and integrate over solid angle, we obtain:

$$\int d\Omega\, P_{l'}(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) V(|\mathbf{x}_1 - \mathbf{x}_2|) = \sum_{l=0}^{\infty} v_l(r_1, r_2) \int d\Omega\, P_{l'}(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) P_l(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) \tag{10.7.2}$$

The angular integral on the right-hand side of (10.7.2) can be performed using the orthogonality of the Legendre polynomials:

$$\begin{aligned} \int d\Omega\, P_{l'}(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) P_l(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) &= 2\pi \int_0^{\pi} d\theta\, \sin\theta P_{l'}(\cos\theta) P_l(\cos\theta) \\ &= \frac{4\pi}{2l+1} \delta_{l',l} \end{aligned} \tag{10.7.3}$$

and we obtain

$$v_l(r_1, r_2) = \frac{2l+1}{4\pi} \int d\Omega\, P_l(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) V(|\mathbf{x}_1 - \mathbf{x}_2|) \tag{10.7.4}$$

10.7.2 Computing the functions $v_l(r_1, r_2)$

We use a method first introduced by Peter Sommer-Larsen Sommer Larsen [Ph.D. Thesis], in which we let the variable r be defined by

$$r := \left(r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta\right)^{1/2} \tag{10.7.5}$$

Then

$$\cos\theta = \frac{r_1^2 + r_2^2 - r^2}{2r_1 r_2} \tag{10.7.6}$$

and

$$\sin\theta\, d\theta = \frac{r}{r_1 r_2}\, dr \tag{10.7.7}$$

whereby we find

$$v_l(r_1, r_2) = \frac{2l+1}{2r_1 r_2} \int_{|r_1-r_2|}^{r_1+r_2} dr\; rV(r) P_l\left(\frac{r_1^2 + r_2^2 - r^2}{2r_1 r_2}\right) \tag{10.7.8}$$

Often the required integrals can be expressed in terms of incomplete Γ-functions. For example when

$$V(r) = r^n e^{-\zeta r} \tag{10.7.9}$$

we have

$$v_l(r_1, r_2) = \frac{2l+1}{2r_1 r_2} \int_{|r_1-r_2|}^{r_1+r_2} dr\; r^{n+1} e^{-\zeta r} P_l\left(\frac{r_1^2 + r_2^2 - r^2}{2r_1 r_2}\right) \tag{10.7.10}$$

Integrals of this type can be evaluated exactly by Mathematica for $l = 0, 1, 2, 3, \ldots$ and $n \geq -1$. The result is

$$v_0(r_1, r_2) = \frac{1}{2\zeta^{n+2} r_1 r_2} \left(\Gamma[n+2, \zeta|r_1 - r_2|] - \Gamma[n+2, \zeta(r_1 + r_2)]\right) \tag{10.7.11}$$

$$\begin{aligned} & v_1(r_1, r_2) \\ = \;& \frac{3}{4\zeta^{n+4} r_1^2 r_2^2} \Big(\zeta^2 (r_1^2 + r_2^2)\{\Gamma[n+2, \zeta|r_1 - r_2|] - \Gamma[n+2, \zeta(r_1 + r_2)]\} \\ & -\Gamma[n+4, \zeta|r_1 - r_2|] + \Gamma[n+4, \zeta(r_1 + r_2)]\Big) \end{aligned} \tag{10.7.12}$$

and so on, where the incomplete Γ-functions are defined by

$$\Gamma[a, z] := \int_z^\infty dt\; t^{a-1} e^{-t} \tag{10.7.13}$$

10.7.3 A Fourier transform solution

Let $F(|\mathbf{x}_1 - \mathbf{x}_2|)$ be any function of the distance $|\mathbf{x}_1 - \mathbf{x}_2|$, and let

$$\begin{aligned} F^t(p) &= \frac{1}{(2\pi)^{3/2}} \int d^3x\; e^{-i\mathbf{p}\cdot\mathbf{x}} F(r) \\ &= \sqrt{\frac{2}{\pi}} \int_0^\infty dr\; r^2 j_0(pr) F(r) \end{aligned} \tag{10.7.14}$$

Then

$$\begin{aligned} F(|\mathbf{x}_1 - \mathbf{x}_2|) &= \frac{1}{(2\pi)^{3/2}} \int d^3p\; e^{i\mathbf{p}\cdot(\mathbf{x}_1 - \mathbf{x}_2)} F^t(p) \\ &= \frac{1}{(2\pi)^{3/2}} \sum_l i^l (2l+1) \int_0^\infty dp\; p^2 j_l(pr_1) F^t(p) \\ &\quad \times \sum_{l'} (-i)^{l'} (2l'+1) j_{l'}(pr_2) \int d\Omega_p\; P_{l'}(\hat{\mathbf{p}}\cdot\hat{\mathbf{x}}_2) P_l(\hat{\mathbf{p}}\cdot\hat{\mathbf{x}}_2) \\ &= \sum_l P_l(\hat{\mathbf{x}}_1\cdot\hat{\mathbf{x}}_2)(2l+1) \sqrt{\frac{2}{\pi}} \int_0^\infty dp\; p^2 j_l(pr_1) j_l(pr_2) F^t(p) \end{aligned} \tag{10.7.15}$$

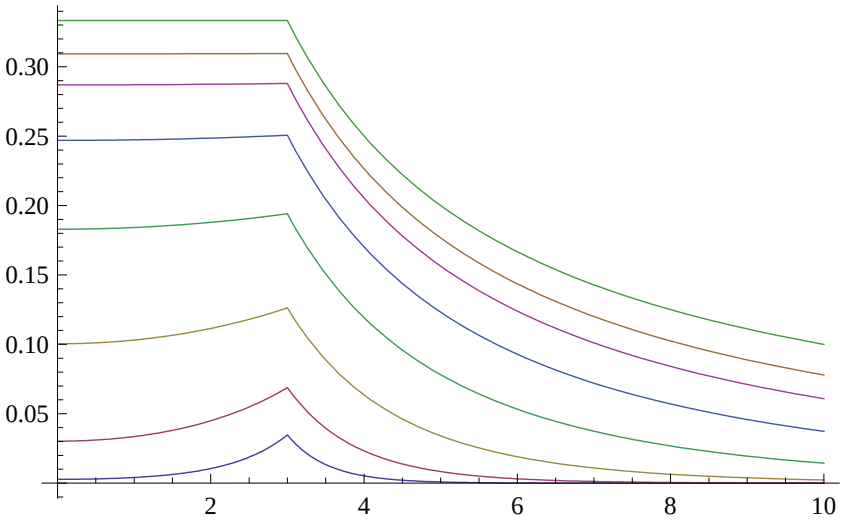


Figure 10.3: The function $v_l(r_1, r_2) = (2l+1)\zeta\, i_l(\zeta r_<) k_l(\zeta r_<)$ compared with $r_<^l / r_>^{l+1}$ for $l = 0$, $r_2 = 3$ and for various values of the screening constant ζ. As $\zeta \to 0$, $v_0(r_1, r_2)$ approaches $1/r_>$ (top curve). The values of ζ used were .025, .05, .1, .2, .4, .8, and 1.6, with 1.6 corresponding to the bottom curve.

Thus we can write

$$F(|\mathbf{x}_1 - \mathbf{x}_2|) \quad = \quad \sum_l f_l(r_1, r_2) P_l(\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2) \tag{10.7.16}$$

where

$$f_l(r_1, r_2) = (2l+1)\sqrt{\frac{2}{\pi}} \int_0^\infty dp\; p^2 j_l(pr_1) j_l(pr_2) F^t(p) \tag{10.7.17}$$

We will discuss later how the methods discussed here for expanding displaced functions in terms of Legendre polynomials may be applied to the problem of evaluating many-center interelectron repulsion integrals when Coulomb Sturmians are used as basis functions for molecular calculations.

Table 10.2 shows the relationships discussed above:

10.7.4 Displacement of functions that do not have spherical symmetry

It might be asked whether the method of Legendre polynomial expansions can be used to represent displaced functions that do not have spherical symmetry. For example, consider the function

$$F(r) x_1^{n_1} x_2^{n_2} x_3^{n_3} \tag{10.7.18}$$

When displaced from a point $\mathbf{B}$ to the origin, this function can be represented as

$$\begin{aligned} &(x_1 - B_1)^{n_1} (x_2 - B_2)^{n_2} (x_3 - B_3)^{n_3} \sum_l f_l(r, B) P_l(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) \\ &= \frac{1}{n!} \frac{\partial^n}{\partial A_1^{n_1} \partial A_2^{n_2} \partial A_3^{n_3}} [\mathbf{A} \cdot (\mathbf{x} - \mathbf{B})]^n \sum_l f_l(r, B) P_l(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) \end{aligned} \tag{10.7.19}$$

where $n = n_1 + n_2 + n_3$. Angular integrals involving functions of the form

$$[\mathbf{A} \cdot (\mathbf{x} - \mathbf{B})]^n P_l(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) \tag{10.7.20}$$

Table 10.2: This table shows the radial functions $f_l(r_<, r_>)$ that appear in Legendre polynomial expansion of a displaced function $F(r)$.

$F(r)$	$F^t(p)$	$f_l(r_<, r_>) = (2l+1)\sqrt{\frac{2}{\pi}}\int_0^\infty dp\, p^2 j_l(pr_1) j_l(pr_2) F^t(p)$
$\frac{1}{r}$	$\sqrt{\frac{2}{\pi}}\frac{1}{p^2}$	$\frac{r_<^l}{r_>^{l+1}}$
$\frac{e^{-\zeta r}}{r}$	$\sqrt{\frac{2}{\pi}}\frac{1}{p^2+\zeta^2}$	$(2l+1)\zeta\, i_l(\zeta r_<) k_l(\zeta r_>)$
$e^{-\zeta r}$	$-\frac{\partial}{\partial\zeta}\sqrt{\frac{2}{\pi}}\frac{1}{p^2+\zeta^2}$	$-\frac{\partial}{\partial\zeta}[(2l+1)\zeta\, i_l(\zeta r_<) k_l(\zeta r_>)]$
$re^{-\zeta r}$	$\frac{\partial^2}{\partial\zeta^2}\sqrt{\frac{2}{\pi}}\frac{1}{p^2+\zeta^2}$	$\frac{\partial^2}{\partial\zeta^2}[(2l+1)\zeta\, i_l(\zeta r_<) k_l(\zeta r_>)]$
$r^2e^{-\zeta r}$	$-\frac{\partial^3}{\partial\zeta^3}\sqrt{\frac{2}{\pi}}\frac{1}{p^2+\zeta^2}$	$-\frac{\partial^3}{\partial\zeta^3}[(2l+1)\zeta\, i_l(\zeta r_<) k_l(\zeta r_>)]$

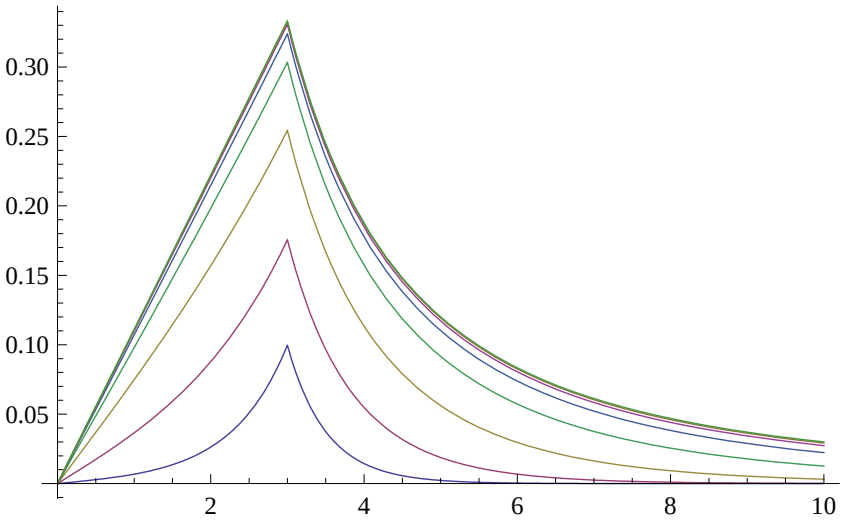


Figure 10.4: This figure is the same as Figure 10.3, except that it shows $v_1(r_1, r_2)$ compared with $r_</r_>^2$ for $r_2 = 3$. The functions $v_l(r_1, r_2)$ occur in the expansion of a screened Coulomb potential in terms of Legendre polynomials. As ζ becomes small, v_1 approaches $r_</r_>^2$.

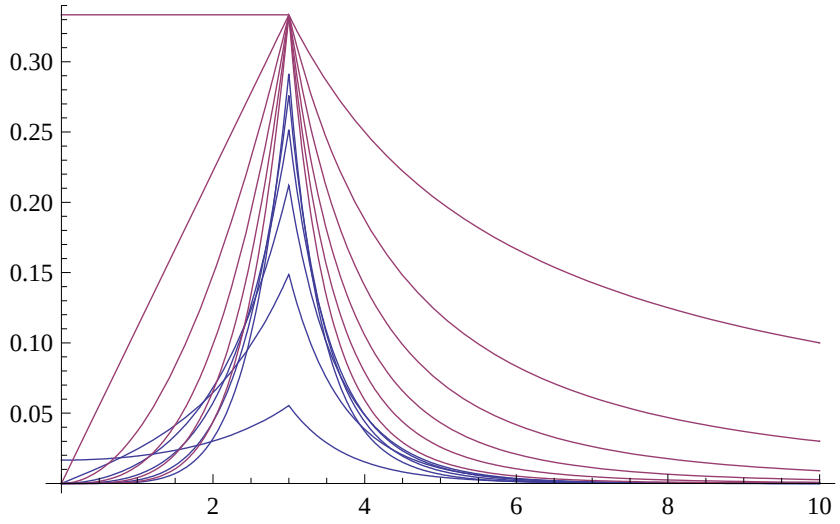


Figure 10.5: This figure shows $v_l(r_1, r_2)$ compared with $r_<^l / r_>^{l+1}$ for $r_2 = 3$ for $\zeta = 1$ and various values of l. For very large values of l, $v_l(r_1, r_2)$ approaches $r_<^l / r_>^{l+1}$.

can easily be evaluated using the vector pairing techniques discussed in Section 10.3.8 as well as in references Avery and Ørmen [1980] and Michels [1981]. In Chapter 5, the methods discussed here are applied to the evaluation of many-center interelectron repulsion integrals for Exponential-Type Orbitals (ETO's).

CHAPTER 11

Evaluation using four-dimensional hyperspherical harmonics

In the methods that will be discussed for evaluating the interelectron repulsion integrals for molecular Sturmians, the hyperspherical method, which will be discussed in this chapter, can be used to rapidly obtain exact results for those integrals that only involve one-center densities, i.e. integrals of the form

$$J = \int d^3x' \, d^3x \, \chi^*_{\mu_1}(\mathbf{x} - \mathbf{X}_a)\chi_{\mu_2}(\mathbf{x} - \mathbf{X}_a)\frac{1}{|\mathbf{x} - \mathbf{x}'|}\chi^*_{\mu_3}(\mathbf{x}' - \mathbf{X}_b)\chi_{\mu_4}(\mathbf{x}' - \mathbf{X}_b) \tag{11.0.1}$$

For the case of integrals involving two-center densities, we expand the displaced Coulomb Sturmians using Legendre-polynomials. This is discussed in Section 11.4. In our pilot calculations, the accuracy of the results using Legendre-polynomial expansion for two-center densities has not been satisfactory. We believe that this is because the accurate representation of a displaced function requires higher terms than we have used in our pilot calculations.

However, the problem of treating two-center molecular Sturmian densities in an accurate and rapid way has been solved by using a special Gaussian expansion method, which will be described in Chapter 12. Because of an automatic scaling property of the molecular Sturmians, this can be done using coefficients that are universal and have been determined once and for all. Much of the calculation can be done off-line, and therefore we can use a greater number of Gaussians to obtain much greater accuracy than is practical in conventional calculations using Gaussians. In addition, Gaussian expansions are used *only* in interelectron repulsion integrals with two-center densities, and all other integrals are carried out exactly. The rapid convergence to zero for the two-center density interelectron integrals alleviates the problem of the wrong asymptotic behavior of Gaussians.

We now present the hyperspherical method, which is exact and rapid in the case of integrals involving only one-center densities.

11.1 Interelectron repulsion integrals for molecular Sturmians

If $\rho_1(\mathbf{x} - \mathbf{X}_a)$ and $\rho_2(\mathbf{x}' - \mathbf{X}_{a'})$ are two electron density distributions, centered respectively on nuclei at the positions $\mathbf{X}_a$ and $\mathbf{X}_{a'}$, the interelectron repulsion between them is given by the integral:

$$J = \int d^3x \int d^3x' \rho_1(\mathbf{x} - \mathbf{X}_a)\frac{1}{|\mathbf{x} - \mathbf{x}'|}\rho_2(\mathbf{x}' - \mathbf{X}_{a'}) \tag{11.1.1}$$

For Coulomb Sturmians, we can rewrite Equation (11.1.1) in terms of the exponent-scaled coordinates $\mathbf{s} = k\mathbf{x}$, $\mathbf{s}' = k\mathbf{x}'$, $\mathbf{S}_a = k\mathbf{X}_a$:

$$J = k \int d^3s \int d^3s' \rho_1(\mathbf{s} - \mathbf{S}_a) \frac{1}{|\mathbf{s} - \mathbf{s}'|} \rho_2(\mathbf{s}' - \mathbf{S}_{a'}) \tag{11.1.2}$$

Thus we see, that for Coulomb Sturmians, the interelectron repulsion integral is equal to the scaling factor k times a pure function of the scaled distances. If we introduce the Fourier transform representation

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{2\pi^2} \int d^3p \, \frac{1}{p^2} e^{-i\mathbf{p}\cdot(\mathbf{x}-\mathbf{x}')} \tag{11.1.3}$$

which in the k-scaled coordinates is

$$\frac{k}{|\mathbf{s} - \mathbf{s}'|} = \frac{k}{2\pi^2} \int d^3q \, \frac{1}{q^2} e^{-i\mathbf{q}\cdot(\mathbf{s}-\mathbf{s}')} \tag{11.1.4}$$

where $\mathbf{q} := \mathbf{p}/k$. We can now rewrite J in the form

$$J = 4\pi k \int d^3q \, \frac{1}{q^2} e^{i\mathbf{q}\cdot\mathbf{S}} \rho_1^t(\mathbf{q}) \rho_2^t(-\mathbf{q}) \tag{11.1.5}$$

where $\mathbf{S} = k(\mathbf{X}_{a'} - \mathbf{X}_a)$ and

$$\rho_j^t(\mathbf{q}) = \frac{1}{(2\pi)^{3/2}} \int d^3s \, \rho_j(\mathbf{s}) e^{-i\mathbf{q}\cdot\mathbf{s}} \qquad j = 1, 2 \tag{11.1.6}$$

11.2 Expansion of products of Coulomb Sturmians

Now let $R_{n,l}(2s)$ be a Coulomb Sturmian radial function with s replaced by $2s$. Then if we make the expansion

$$g(s) = \sum_n a_n R_{n,l}(2s) \tag{11.2.1}$$

then from the potential-weighted orthonormality relations (10.1.6) it follows that the expansion coefficients will be given by

$$a_n = \frac{n}{2} \int_0^\infty ds \; s \; R_{n,l}(2s) g(s) \tag{11.2.2}$$

We next consider the density

$$\rho_{\mu',\mu}(\mathbf{s}) = \chi^*_{\mu'}(\mathbf{s}) \chi_\mu(\mathbf{s}) = R_{n',l'}(s) R_{n,l}(s) Y^*_{l',m'}(\hat{\mathbf{s}}) Y_{l,m}(\hat{\mathbf{s}}) \tag{11.2.3}$$

If we now make the expansion

$$\rho_{\mu',\mu}(\mathbf{s}) = \sum_{\mu''} R_{n'',l''}(2s) Y_{l'',m''}(\hat{\mathbf{s}}) C_{\mu'';\mu',\mu} := \sum_{\mu''} \chi_{\mu''}(2\mathbf{s}) C_{\mu'';\mu',\mu} \tag{11.2.4}$$

then the coefficients in the expansion will be given by

$$\begin{aligned} C_{\mu'';\mu',\mu} \quad = \quad & \frac{n''}{2} \int_0^\infty ds \; s \; R_{n'',l''}(2s) R_{n',l'}(s) R_{n,l}(s) \\ & \times \int d\Omega_3 \; Y^*_{l'',m''}(\hat{\mathbf{s}}) Y^*_{l',m'}(\hat{\mathbf{s}}) Y_{l,m}(\hat{\mathbf{s}}) \end{aligned} \tag{11.2.5}$$

Like the coefficients $c_{\mu'';\mu',\mu}$ defined in Equation (10.5.21), the coefficients $C_{\mu'';\mu',\mu}$ form a large but very sparse matrix that can be precalculated once and for all and stored.

11.3 Interelectron repulsion integrals from hyperspherical harmonics

We would like to evaluate the integral

$$J_{\mu',\mu} = 4\pi k \int d^3q \frac{1}{q^2} e^{i\mathbf{q}\cdot\mathbf{S}} \rho^t_{\mu'}(\mathbf{q}) \rho^t_{\mu}(-\mathbf{q}) \tag{11.3.1}$$

where

$$\begin{aligned} \rho_{\mu'}(\mathbf{s}) &= R_{n',l'}(2s) Y_{l',m'}(\hat{\mathbf{s}}) \\ \rho_{\mu}(\mathbf{s}) &= R_{n,l}(2s) Y_{l,m}(\hat{\mathbf{s}}) \end{aligned} \tag{11.3.2}$$

so that

$$\begin{aligned} \rho^t_{\mu'}(\mathbf{q}) &= M(q) Y_{n'-1,l',m'}(\hat{\mathbf{u}}) := M(q) Y_{\mu'}(\hat{\mathbf{u}}) \\ \rho^t_{\mu}(-\mathbf{q}) &= (-1)^l M(q) Y_{n-1,l,m}(\hat{\mathbf{u}}) := (-1)^l M(q) Y_{\mu}(\hat{\mathbf{u}}) \end{aligned} \tag{11.3.3}$$

always remembering that k=2 should be used in $M(q)$ and in Fock's expressions for the $\mathbf{u}$'s. Then

$$J_{\mu',\mu} = 4\pi k \int d^3q \frac{1}{q^2} e^{i\mathbf{q}\cdot\mathbf{S}} M^2(q) (-1)^l Y_{\mu'}(\mathbf{u}) Y_{\mu}(\mathbf{u}) \tag{11.3.4}$$

Since we know how to represent the product $(-1)^l Y_{\mu'}(\mathbf{u}) Y_{\mu}(\mathbf{u})$ in terms of $Y_{\mu''}(\mathbf{u})$, we can express the matrix $J_{\mu',\mu}$ in terms of a single vector, $J_{\mu''}$: Let

$$\tilde{c}_{\mu'';\mu',\mu} := (-1)^l \int d\Omega_4 Y^*_{\mu''}(\mathbf{u}) Y_{\mu'}(\mathbf{u}) Y_{\mu}(\mathbf{u}) \tag{11.3.5}$$

Then

$$J_{\mu',\mu} = \sum_{\mu''} J_{\mu''} \tilde{c}_{\mu'';\mu',\mu} \tag{11.3.6}$$

where

$$J_{\mu} := 4\pi k \int d^3q \frac{1}{q^2} e^{i\mathbf{q}\cdot\mathbf{S}} M^2(q) Y_{\mu}(\mathbf{u}) \tag{11.3.7}$$

The sum in Equation (11.3.6) is a small, terminating sum with very few nonzero terms. The coefficients $\tilde{c}_{\mu'';\mu',\mu}$ differ slightly from the coefficients $c_{\mu'';\mu',\mu}$, but they too form a large but very sparse matrix that can be precalculated and stored. We must now evaluate J_{μ}. To do so, it is convenient to introduce the notation

$$\rho^t_{\mu}(\mathbf{q}) = M(q) Y_{\mu}(\mathbf{u}) := i^{-l} f_{n,l}(q) Y_{l,m}(\hat{\mathbf{q}}) \tag{11.3.8}$$

Then

$$\begin{aligned} J_{\mu} &= 4\pi k \int d^3q \, \frac{1}{q^2} e^{i\mathbf{q}\cdot\mathbf{S}} M^2(q) Y_{\mu}(\mathbf{u}) \\ &= 4\pi k i^{-l} \int d^3q \, \frac{1}{q^2} e^{i\mathbf{q}\cdot\mathbf{S}} M(q) f_{n,l}(q) Y_{l,m}(\hat{\mathbf{q}}) \\ &= 4\pi k i^{-l} \int_0^\infty dq \, M(q) f_{n,l}(q) \int d\Omega_p \, e^{i\mathbf{q}\cdot\mathbf{S}} Y_{l,m}(\hat{\mathbf{q}}) \\ &= (4\pi)^2 k i^{-l} \sum_{l'=0}^{\infty} i^{l'} \sum_{m'=-l'}^{l'} Y_{l',m'}(\hat{\mathbf{S}}) \\ &\quad \times \int_0^\infty dq \, M(q) f_{n,l}(q) j_l(qS) \int d\Omega_p \, Y^*_{l',m'}(\hat{\mathbf{q}}) Y_{l,m}(\hat{\mathbf{q}}) \\ &= Y_{l,m}(\hat{\mathbf{S}}) \, (4\pi)^2 \, k \int_0^\infty dq \, M(q) f_{n,l}(q) j_l(qS) \end{aligned} \tag{11.3.9}$$

The radial q-integrals in equation (11.3.8) are simple enough to be evaluated exactly by Mathematica, and they depend only on n and l. We are now in a position to evaluate interelectron repulsion integrals of the form

$$J_{\mu_1,\mu_2,\mu_3,\mu_4} = k \int d^3s \int d^3s' \chi^*_{\mu_1}(\mathbf{s}-\mathbf{S}_a)\chi_{\mu_2}(\mathbf{s}-\mathbf{S}_a)\frac{1}{|\mathbf{s}-\mathbf{s}'|}\chi^*_{\mu_3}(\mathbf{s}'-\mathbf{S}_{a'})\chi_{\mu_4}(\mathbf{s}'-\mathbf{S}_{a'}) \quad (11.3.10)$$

Making use of equation (11.2.3) and (12.1.11), we have

$$\begin{aligned} J_{\mu',\mu} &= \sum_{\mu''} J_{\mu''}\tilde{c}_{\mu'';\mu',\mu} \\ J_{\mu_1,\mu_2,\mu_3,\mu_4} &= \sum_{\mu',\mu} J_{\mu',\mu} C_{\mu';\mu_1,\mu_2} C_{\mu;\mu_3,\mu_4} \end{aligned} \quad (11.3.11)$$

The sums in Equation (11.3.11) terminate and are small, with few non-zero terms. Because of the sparseness of the coefficient matrices, the sums in (11.3.10) can be performed very rapidly. We will see in the next section, that when expansions of two-center densities are used, the analogous equation involves an infinite series of coefficients obtained through the Legendre-expansion method, giving rise to an infinite sum that must be truncated.

For diatomic molecules, the integrals $J_{\mu_1,\mu_2,\mu_3,\mu_4}$ can be shown to be pure functions of $S = kR$, provided that the coordinate axes are taken along the internuclear axis. The first few of these are

$$\begin{aligned} J_{1111} &= \frac{e^{-2S}(24+(-24-S(33+2S(9+2S))))}{24S} \\ J_{1122} &= \frac{e^{-2S}(-240+240e^{-2S}-390S-300S^2-145S^3-50S^4-12S^5)}{240S} \\ J_{1221} &= \frac{e^{-2S}(-120+120e^{2S}-195S-150S^2-80S^3-40S^4-16S^5)}{480S} \end{aligned} \quad (11.3.12)$$

where the four subscripts to J refer to the indices into the list

$$(n,l,m) \in \{(1,0,0),(2,0,0),(2,1,-1),(2,1,0),(2,1,1),\ldots\} \quad (11.3.13)$$

The integrals

$$I_{n,l}(S) := \int_0^\infty dq\; M(q) f_{n,l}(q) j_l(qS) \quad (11.3.14)$$

can be calculated once and for all and stored. These integrals are few in number, since they depend only on n and l and are independent of m. The first 105 of these functions, corresponding to $n_{\max} = 14$, are shown in Figures (11.1) and (11.2).

11.4 Many-center integrals treated by the method of Legendre polynomial expansions

Now let us consider the case where we wish to evaluate 3-center or 4-center interelectron repulsion integrals. Then we must evaluate

$$J = 4\pi \int d^3p \frac{1}{p^2} e^{i\mathbf{p}\cdot\mathbf{A}} \rho^t(\mathbf{p})\rho'^t(-\mathbf{p}) \quad (11.4.1)$$

where either the density $\rho(\mathbf{x})$, $\rho'(\mathbf{x})$, or both may be due to orbitals located on two different centers. This is a more difficult case than that of the single-center densities discussed above,

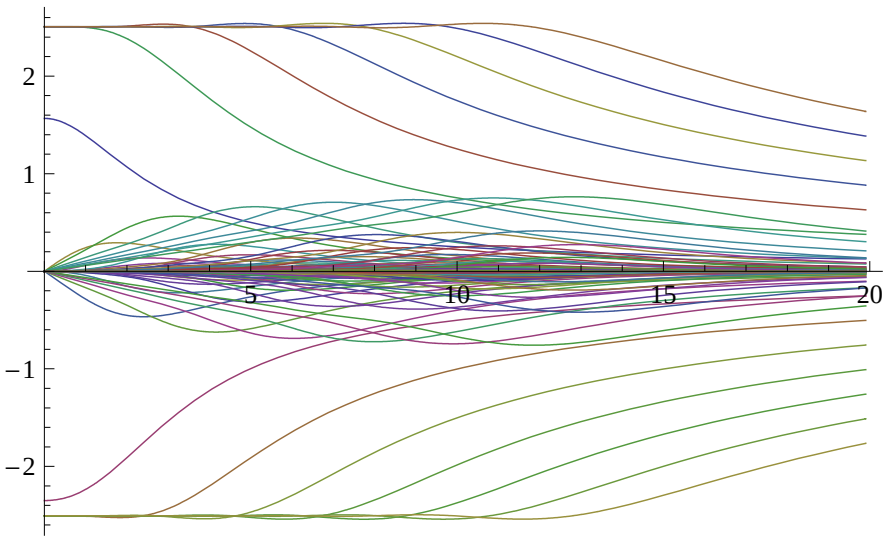


Figure 11.1: The integrals $I_{n,l}(R)$ of Equation (11.3.14) are shown here plotted as functions of R. There are 105 functions, corresponding $n = 1, 2, \ldots, 14$ and $l = 0, 1, \ldots, n-1$.

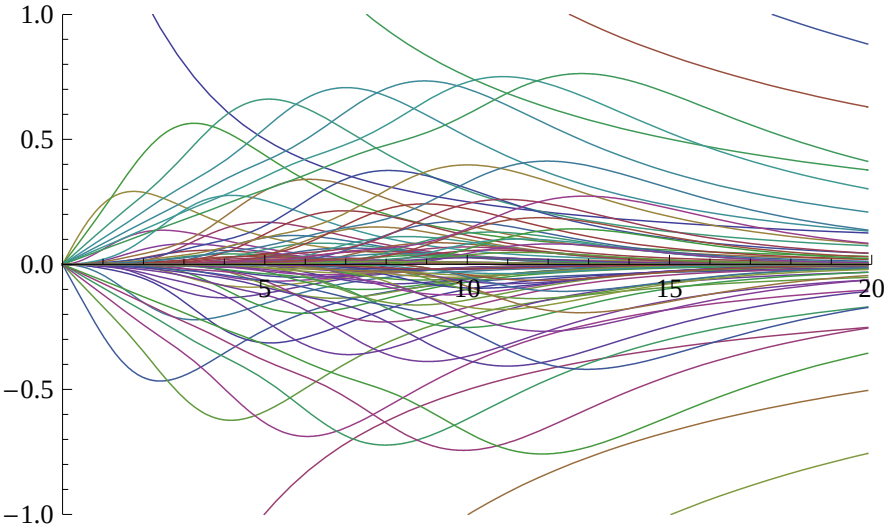


Figure 11.2: The functions $I_{n,l}(R)$ shown in more detail. For small values of R the integrals are proportional to R^l, while for large values they are proportional to $1/R^{l+1}$.

but nevertheless it is always possible to make an expansion in terms of Coulomb Sturmian basis functions, since these form a complete set. Thus we write

$$\begin{aligned} \rho(\mathbf{x}) &= \sum_{\mu} a_{\mu}\chi_{\mu}(\mathbf{x}) \\ \rho'(\mathbf{x}) &= \sum_{\mu'} a'_{\mu'}\chi_{\mu'}(\mathbf{x}) \end{aligned} \tag{11.4.2}$$

where the χ's are Coulomb Sturmian basis functions with double the k value. It is always possible to do this, whether the ρ's are 1-center densities or 2-center densities since the χ's span $H^1(\mathbb{R}^3)$. Then

$$\begin{aligned} \rho^t_{\mu}(\mathbf{p}) &= \sum_{\mu} a_{\mu} M(p) Y_{\mu}(\hat{\mathbf{u}}) \\ \rho'^t_{\mu'}(-\mathbf{p}) &= \sum_{\mu'} a_{\mu'} (-1)^{l'} M(p) Y_{\mu'}(\hat{\mathbf{u}}) \end{aligned} \tag{11.4.3}$$

and

$$\begin{aligned} J &= 4\pi \sum_{\mu',\mu} a'_{\mu'} a_{\mu} \int d^3p \frac{1}{p^2} e^{i\mathbf{p}\cdot\mathbf{S}} M^2(p) (-1)^l Y_{\mu'}(\hat{\mathbf{u}}) Y_{\mu}(\hat{\mathbf{u}}) \\ &:= \sum_{\mu',\mu} a'_{\mu'} a_{\mu} J_{\mu',\mu} \end{aligned} \tag{11.4.4}$$

But the integrals $J_{\mu',\mu}$ are just the familiar ones that we are able to evaluate with ease and store, as was discussed above. Thus almost the whole computational weight of evaluating J lies in the problem of making expansions of the form shown in equation (11.4.2).

If we write $\rho(\mathbf{x})$ as

$$\rho(\mathbf{x}) = \sum_{j} f_j(r) F_j(\hat{\mathbf{x}}) \tag{11.4.5}$$

and if

$$\chi_{\mu}(\mathbf{x}) = R_{n,l}(2kr) Y_{l,m}(\hat{\mathbf{x}}) \tag{11.4.6}$$

then

$$\begin{aligned} \int d^3x \chi^*_{\mu'}(\mathbf{x}) \frac{1}{r} \rho(\mathbf{x}) &= \sum_{\mu} a_{\mu} \int d^3x \chi^*_{\mu'}(\mathbf{x}) \frac{1}{r} \chi_{\mu}(\mathbf{x}) \\ &= \sum_{\mu} a_{\mu} \delta_{\mu',\mu} \frac{2k}{n} \end{aligned} \tag{11.4.7}$$

and

$$\begin{aligned} a_{\mu} &= \frac{n}{2k} \int d^3x \; \chi^*_{\mu}(\mathbf{x}) \frac{1}{r} \rho(\mathbf{x}) \\ &= \frac{n}{2k} \int d^3x \; R_{n,l}(2kr) Y^*_{l,m}(\hat{\mathbf{x}}) \frac{1}{r} \sum_j f_j(r) F_j(\hat{\mathbf{x}}) \\ &= \sum_j \frac{n}{2k} \int_0^{\infty} dr \; r \; R_{n,l}(2kr) f_j(r) \int d\Omega \; Y^*_{l,m}(\hat{\mathbf{x}}) F_j(\hat{\mathbf{x}}) \end{aligned} \tag{11.4.8}$$

The angular integral can be evaluated exactly in all cases. For two-center densities, the radial integral in (11.4.8) is

$$\begin{aligned} &\frac{n}{2k} \int_0^{\infty} dr \; r \; R_{n,l}(2kr) f_j(r) \\ = \; &\frac{n}{2k} \int_0^{\infty} dr \; r \; R_{n,l}(2kr) R_{n',l'}(kr) v_{l''}(r, C) \end{aligned} \tag{11.4.9}$$

Thus the fundamental radial integral has the form

$$\int_0^\infty dr\, r^{n'+1} e^{-3kr} v_l[r^n e^{-kr}](r, C) \tag{11.4.10}$$

In equation (11.4.10), the expression in square brackets indicates the function that is being moved from the point **C** to the origin.

Let us now turn to the problem of evaluating the angular integrals in equation (11.4.8). To do this, we make use of the methods discussed in Sections 10.3 and 10.7. From Section 10.7, we know that if $V(r)$ is any spherically symmetric function, the displaced function $V(|\mathbf{x}-\mathbf{B}|)$ can be expanded in a series of the form

$$V(|\mathbf{x}-\mathbf{B}|) = \sum_l v_l(r, B) P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}}) \tag{11.4.11}$$

Now let us consider a function of the form $V(r)x_1$. The corresponding displaced function will then be given by

$$V(|\mathbf{x}-\mathbf{B}|)(x_1 - B_1) = \sum_l v_l(r, B) P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})(x_1 - B_1) \tag{11.4.12}$$

But

$$(x_1 - B_1) = \frac{\partial}{\partial A_1}(\mathbf{x}-\mathbf{B})\cdot\hat{\mathbf{A}} \tag{11.4.13}$$

and more generally

$$\begin{aligned}
&(x_1 - B_1)^{n_1}(x_2 - B_2)^{n_2}(x_3 - B_3)^{n_3} \\
&= \frac{1}{n!}\frac{\partial^n}{\partial A_1^{n_1} A_2^{n_2} A_3^{n_3}}[(\mathbf{x}-\hat{\mathbf{B}})\cdot\hat{\mathbf{A}}]^n \\
&= \frac{1}{n!}\frac{\partial^n}{\partial A_1^{n_1} A_2^{n_2} A_3^{n_3}}(\mathbf{x}\cdot\hat{\mathbf{A}} - \hat{\mathbf{A}}\cdot\mathbf{B})^n \qquad n := n_1 + n_2 + n_3
\end{aligned} \tag{11.4.14}$$

Thus the form

$$(\mathbf{x}\cdot\hat{\mathbf{A}} - \hat{\mathbf{A}}\cdot\mathbf{B})^n \sum_l v_l(r, B) P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}}) \tag{11.4.15}$$

which represents $(\mathbf{x}\cdot\hat{\mathbf{A}})^n V(r)$ displaced to the origin from the point **B**, is sufficient to represent any displaced function, regardless of its angular dependence. The angular dependence appropriate to particular situations can be found by partial differentiation with respect to the components of $\hat{\mathbf{A}}$. We now remember that the two-center density $\rho(\mathbf{x})$ is composed of an orbital located at the origin as well as the orbital located at **B**. If the orbital located at the origin has the angular dependence $Y_{l',m'}(\hat{\mathbf{x}})$ then the angular integrals in (11.4.8) have the form

$$\int d\Omega\, Y^*_{l,m}(\hat{\mathbf{x}}) Y_{l',-m}(\hat{\mathbf{x}})(\mathbf{x}\cdot\hat{\mathbf{A}} - \hat{\mathbf{A}}\cdot\mathbf{B})^n P_{l''}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}}) \tag{11.4.16}$$

The expression $(\mathbf{x}\cdot\hat{\mathbf{A}} - \hat{\mathbf{A}}\cdot\mathbf{B})^n$ can be expanded using the binomial theorem:

$$\begin{aligned}
(\mathbf{x}\cdot\hat{\mathbf{A}} - \hat{\mathbf{A}}\cdot\mathbf{B})^n &= (\mathbf{x}\cdot\hat{\mathbf{A}})^n - \frac{1}{2!}(\hat{\mathbf{A}}\cdot\mathbf{B})(\mathbf{x}\cdot\hat{\mathbf{A}})^{n-1} + \cdots + (\hat{\mathbf{A}}\cdot\mathbf{B})^n \\
&= r^n(\hat{\mathbf{x}}\cdot\hat{\mathbf{A}})^n - \frac{r^{n-1}B}{2!}(\hat{\mathbf{A}}\cdot\hat{\mathbf{B}})(\hat{\mathbf{x}}\cdot\hat{\mathbf{A}})^{n-1} + \cdots + r^0(\hat{\mathbf{A}}\cdot\mathbf{B})^n
\end{aligned} \tag{11.4.17}$$

Consequently, the fundamental angular integral with which we are faced has the form:

$$\int d\Omega\, Y_{l,m}(\hat{\mathbf{x}})(\hat{\mathbf{x}}\cdot\hat{\mathbf{A}})^n P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}}) \tag{11.4.18}$$

This integral can easily be evaluated and stored using the methods of Section 10.3.8.

11.4.1 A simple example

As a simple example, let us consider the case where both the orbital on point **B** and the orbital at the origin are 1s Coulomb Sturmians, so that (recalling that $\chi_{1,0,0}(\mathbf{x}) = R_{10}(kr)Y_{00}(\hat{\mathbf{x}}) = 2k^{3/2}e^{-kr}\cdot\frac{1}{\sqrt{4\pi}} = \sqrt{\frac{k^3}{\pi}}e^{-kr}$)

$$\begin{aligned}\rho(\mathbf{x}) &= \chi^*_{1,0,0}(\mathbf{x})\chi_{1,0,0}(\mathbf{x}-\mathbf{B}) \\ &= \frac{k^3}{\pi}e^{-kr}e^{-k|\mathbf{x}-\mathbf{B}|} \\ &= \frac{k^3}{\pi}e^{-kr}\sum_{l'} v_{l'}(r,B)P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\end{aligned} \tag{11.4.19}$$

Then

$$\begin{aligned}a_{\mu;\mu_1,n_2} &= \frac{n}{2k}\int d^3x\ \chi^*_\mu(\mathbf{x})\frac{1}{r}\rho(\mathbf{x}) \\ &= \frac{n}{2k}\int d^3x\ R_{n,l}(2kr)Y^*_{l,m}(\hat{\mathbf{x}})\frac{1}{r}\frac{k^3}{\pi}e^{-kr}\sum_{l'} v_{l'}(r,B)P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}}) \\ &= \frac{\sqrt{2}k^{7/2}}{\pi}n\int_0^\infty dr\ r\sum_{n'=l}^{n-1} c^{n'}_{n,l}(2kr)^{n'}\ e^{-3kr}\sum_{l'} v_{l'}(r,B)\int d\Omega\ Y^*_{l,m}(\hat{\mathbf{x}})P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\end{aligned} \tag{11.4.20}$$

where $c^{n'}_{nl}$ is the coefficient of $r^{n'}$ in the polynomial part of $R_{nl}(r)$. If **B** is in the direction of the z-axis, the angular integral vanishes unless $m = 0$. However, for general **B**, we have to write

$$a_\mu = \frac{2(2k)^{3/2}n}{2l+1}Y^*_{l,m}(\hat{\mathbf{B}})\sum_{n'=l}^{n-1} c^{n'}_{n,l}2^{n'}\int_0^\infty d(kr)\ (kr)^{n'+1}e^{-3kr}v_l[e^{-kr}](r,B) \tag{11.4.21}$$

Here, $v_l[e^{-kr}](r,B)$ is calculated according to Section 10.7. The integral in Equation (11.4.21) can be calculated exactly by Mathematica. All in all, the coefficients a_{nlm} can be precalculated and stored as closed form expressions.

11.4.2 A slightly more general example

Now let us consider a slightly more complicated example, where the two basis functions in the density $\rho(\mathbf{x})$ are respectively n_1s and n_2s Coulomb Sturmians. Then

$$\begin{aligned}\rho(\mathbf{x}) &= \chi^*_{n_1,0,0}(\mathbf{x})\chi_{n_2,0,0}(\mathbf{x}-\mathbf{B}) \\ &= \frac{k^3}{4\pi}\sum_{n_1'=0}^{n_1} c^{n_1'}_{n_1,0}\sum_{n_2'=0}^{n_2} c^{n_2'}_{n_2,0}(kr)^{n_1'}e^{-kr}\sum_{l'} v_{l'}[(kr)^{n_2'}e^{-kr}](r,B)P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\end{aligned} \tag{11.4.22}$$

The new version of equation (11.4.20) becomes (recalling that we now have $\chi_\mu(\mathbf{x}) = R_{nl}(2kr)Y_{lm}(\hat{\mathbf{x}})$)

$$\begin{aligned}a_{\mu,n_1,n_2} &= \frac{n}{2k}\int d^3x\ \chi^*_\mu(\mathbf{x})\frac{1}{r}\rho(\mathbf{x}) \\ &= \frac{n(2k)^{3/2}}{2(2l+1)}Y^*_{l,m}(\hat{\mathbf{B}})\sum_{n_1'=0}^{n_1} c^{n_1'}_{n_1,0}\sum_{n_2'=0}^{n_2} c^{n_2'}_{n_2,0}\sum_{n'=l}^{n-1} c^{n'}_{n,l}2^{n'} \\ &\times\int_0^\infty d(kr)\ (kr)^{n'+n_1'+1}e^{-3kr}v_l[(kr)^{n_2'}e^{-kr}](r,B)\end{aligned} \tag{11.4.23}$$

This result reduces to (11.4.21) in the case where $n_1 = 1$, $n_2 = 1$. Again, we find that we can precalculate the a_{nlm} as closed form expressions that depend only on $s = kr$, times a spherical harmonic. While one might fear that these three-way sums give rise to very large expressions, they in fact can be automatically reduced to moderately simple expressions.

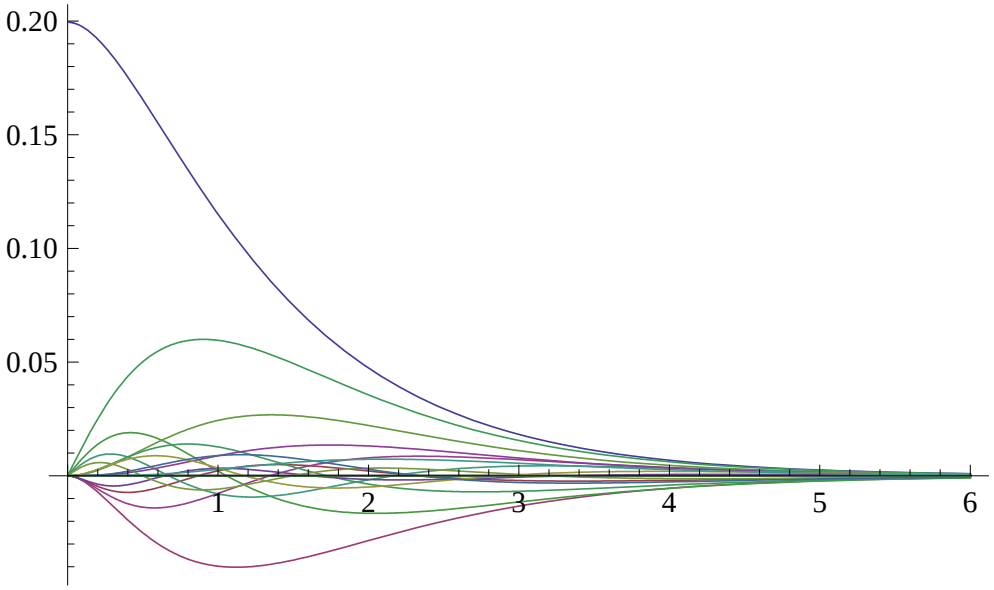


Figure 11.3: The non-zero a's of equation (11.4.21) as functions of B, when $\hat{\mathbf{B}} = (0, 0, 1)$

11.4.3 The next higher degree of difficulty

The next higher degree of difficulty is reached when we consider the case where the orbital at the origin lacks spherical symmetry, but the orbital at $\mathbf{B}$ is still spherically symmetric. Then the 2-center density $\rho(\mathbf{x})$ has the form:

$$\begin{aligned}\rho_{\mu_1,\mu_2}(\mathbf{x}) &= \chi^*_{n_1,l_1,m_1}(\mathbf{x})\chi_{n_2,0,0}(\mathbf{x}-\mathbf{B}) \\ &= \frac{k^3}{\sqrt{4\pi}} \sum_{n_1'=l_1}^{n_1} c^{n_1'}_{n_1,l_1} \sum_{n_2'=0}^{n_2} c^{n_2'}_{n_2,0} Y^*_{l_1,m_1}(\hat{\mathbf{x}})(kr)^{n_1'} e^{-kr} \\ &\times \sum_{l'} v_{l'}[(kr)^{n_2'} e^{-kr}](r,B) P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\end{aligned} \tag{11.4.24}$$

Then we find that

$$\begin{aligned}a_\mu = \frac{nk^2}{4\sqrt{\pi}} \sum_{n_1'=l_1}^{n_1} c^{n_1'}_{n_1,l_1} \sum_{n_2'=0}^{n_2} c^{n_2'}_{n_2,0} \int_0^\infty dr\, r\, R_{n,l}(2kr)(kr)^{n_1'} e^{-kr} \\ \times \sum_{l'} v_{l'}[(kr)^{n_2'} e^{-kr}](r,B) \int d\Omega\, P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}}) Y^*_{l,m}(\hat{\mathbf{x}}) Y^*_{l_1,m_1}(\hat{\mathbf{x}})\end{aligned} \tag{11.4.25}$$

The angular integral in Equation (11.4.25) can easily be computed using the methods given in Section 10.3.8 or Section 10.3.7.

11.4.4 The totally general case

Drunk with power, we go on to the totally general case. Consider the function

$$V(r)x_1^{j_1}x_2^{j_2}x_3^{j_3} \tag{11.4.26}$$

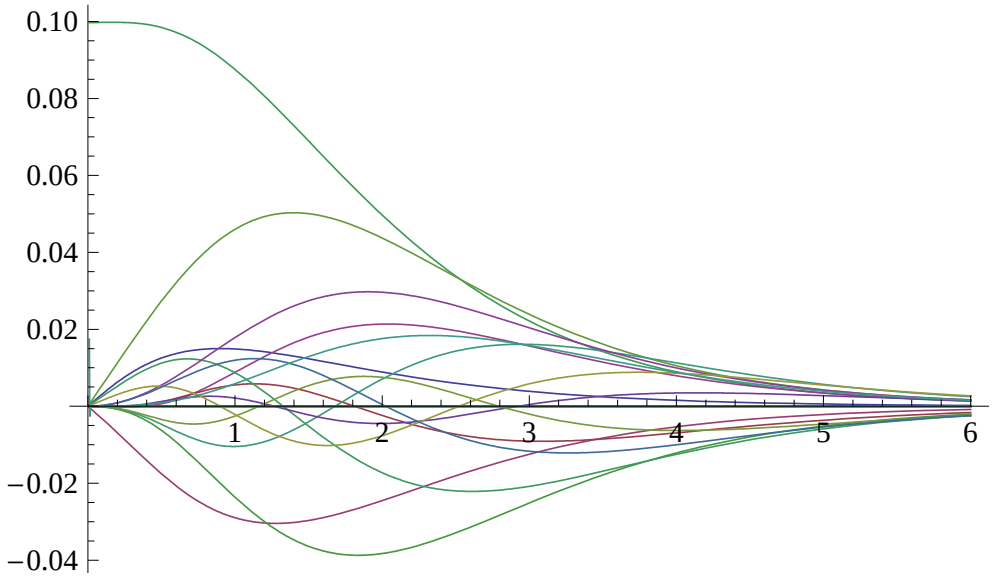


Figure 11.4: The non-zero a's of equation (11.4.25) as functions of B, when $\hat{\mathbf{B}} = (0,0,1)$, $(n_1,l_1,m_1) = (2,1,0)$, and $(n_2,l_2,m_2) = (1,0,0)$.

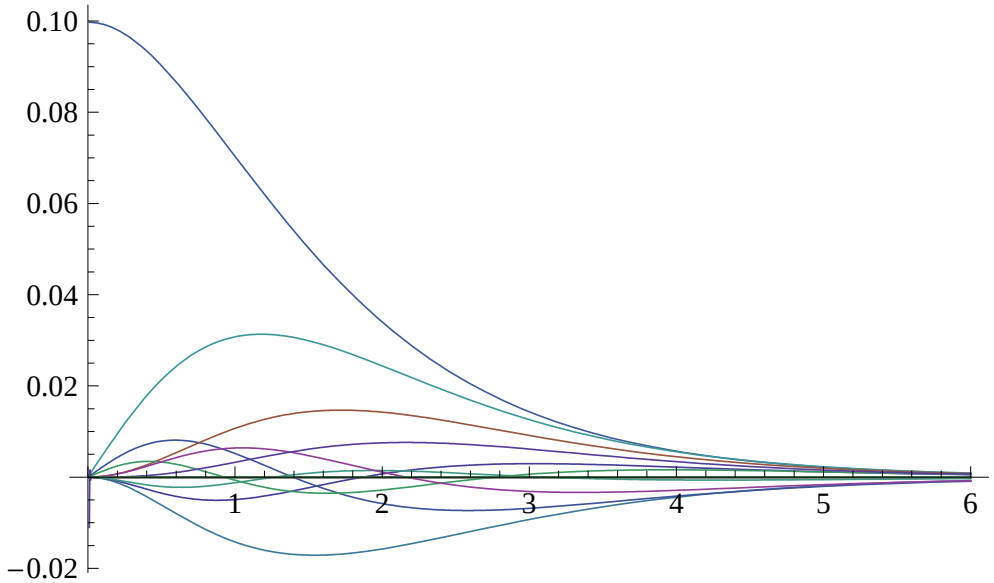


Figure 11.5: The non-zero a's of equation (11.4.25) as functions of B, when $\hat{\mathbf{B}} = (0,0,1)$, $(n_1,l_1,m_1) = (2,1,1)$, and $(n_2,l_2,m_2) = (1,0,0)$.

We saw earlier in this section, that when displaced from a point **B** to the origin, this function can be represented as

$$\begin{aligned}&(x_1 - B_1)^{j_1}(x_2 - B_2)^{j_2}(x_3 - B_3)^{j_3}\sum_l v_l(r,B)P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\\ &= \frac{1}{l_2!}\frac{\partial^{l_2}}{\partial A_1^{j_1}\partial A_2^{j_2}\partial A_3^{j_3}}[\hat{\mathbf{A}}\cdot(\mathbf{x}-\mathbf{B})]^{l_2}\sum_l v_l(r,B)P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\end{aligned} \tag{11.4.27}$$

where $l_2 = j_1 + j_2 + j_3$. Angular integrals involving functions of the form

$$[\hat{\mathbf{A}}\cdot(\mathbf{x}-\mathbf{B})]^{l_2}P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}}) \tag{11.4.28}$$

can easily be evaluated using the vector pairing techniques discussed in Section 10.3. We next consider a function of the form

$$V(r)\sum_{j_1+j_2+j_3=l} y_{\mathbf{j},l_2,m_2}\, x_1^{j_1}x_2^{j_2}x_3^{j_3} = V(r)r^{l_2}Y_{l_2,m_2}(\hat{\mathbf{x}}) \tag{11.4.29}$$

where the coefficients $y_{\mathbf{j},l,m}$ are the coefficients of the harmonic polynomial corresponding to Y_{lm}.

$$Y_{l_2,m_2}(\hat{\mathbf{x}}) = \sum_{j_1+j_2+j_3=l} y_{\mathbf{j},l_2,m_2}\,\hat{x}_1^{j_1}\hat{x}_2^{j_2}\hat{x}_3^{j_3} = \frac{1}{r^{l_2}}\sum_{j_1+j_2+j_3=l} y_{\mathbf{j},l_2,m_2}\, x_1^{j_1}x_2^{j_2}x_3^{j_3} \tag{11.4.30}$$

If we define in a similar way

$$Y_{lm}(\nabla) = \frac{1}{l!}\sum_{j_1+j_2+j_3=l} y_{\mathbf{j},l,m}\,\frac{1}{\partial x_1^{j_1}x_2^{j_2}x_3^{j_3}} \tag{11.4.31}$$

then we can represent the function in (11.4.29) displaced from point **B** to the origin as

$$\begin{aligned}&\frac{1}{l_2!}\sum_{\mathbf{j}} y_{\mathbf{j},l_2,m_2}\frac{\partial^{l_2}}{\partial A_1^{j_1}\partial A_2^{j_2}\partial A_3^{j_3}}[\hat{\mathbf{A}}\cdot(\mathbf{x}-\mathbf{B})]^{l_2}\sum_l v_l[V(r)](r,B)P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\\ &= Y_{l_2m_2}(\nabla_A)\left([\hat{\mathbf{A}}\cdot(r\hat{\mathbf{x}}-B\hat{\mathbf{B}})]^{l_2}\sum_l v_l[V(r)](r,B)P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\right)\\ &= Y_{l_2m_2}(\nabla_A)\sum_{q=0}^{l_2}\binom{l_2}{q}B^{l_2-q}(\hat{\mathbf{A}}\cdot\hat{\mathbf{B}})^{l_2-q}\sum_l r^q v_l[V(r)](r,B)(\hat{\mathbf{x}}\cdot\hat{\mathbf{A}})^q P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\end{aligned} \tag{11.4.32}$$

Finally, let us consider the Coulomb Sturmian basis function $\chi_{n_2l_2m_2}(\mathbf{x})$. When displaced from the point **B** to the origin, it can be represented as

$$\begin{aligned}\chi_{n_2,l_2,m_2}(\mathbf{x}-\mathbf{B}) = k^{3/2}&\sum_{n_2'=l_2}^{n_2} c_{n_2,l_2}^{n_2'}Y_{l_2m_2}(\nabla_A)\sum_{q=0}^{l_2}\binom{l_2}{q}(kB)^{l_2-q}(\hat{\mathbf{A}}\cdot\hat{\mathbf{B}})^{l_2-q}\\ &\times\sum_l v_l[(kr)^{q+n_2'-l_2}e^{-kr}](r,B)(\hat{\mathbf{x}}\cdot\hat{\mathbf{A}})^q P_l(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})\end{aligned} \tag{11.4.33}$$

Then the 2-center density is given by

$$\begin{aligned}\rho(\mathbf{x}) &= \chi_{\mu_1}^*(\mathbf{x})\chi_{\mu_2}(\mathbf{x}-\mathbf{B})\\ &= k^3\sum_{n_1'=l_1}^{n_1} c_{n_1,l_1}^{n_1'}\sum_{n_2'=l_2}^{n_2} c_{n_2,l_2}^{n_2'}Y_{l_2m_2}(\nabla_A)\sum_{q=0}^{l_2}\binom{l_2}{q}(kB)^{l_2-q}(\hat{\mathbf{A}}\cdot\hat{\mathbf{B}})^{l_2-q}\\ &\times(kr)^{n_1'}e^{-kr}\sum_{l'} v_{l'}[(kr)^{q+n_2'-l_2}e^{-kr}](r,B)(\hat{\mathbf{x}}\cdot\hat{\mathbf{A}})^q P_{l'}(\hat{\mathbf{x}}\cdot\hat{\mathbf{B}})Y_{l_1,m_1}^*(\hat{\mathbf{x}})\end{aligned} \tag{11.4.34}$$

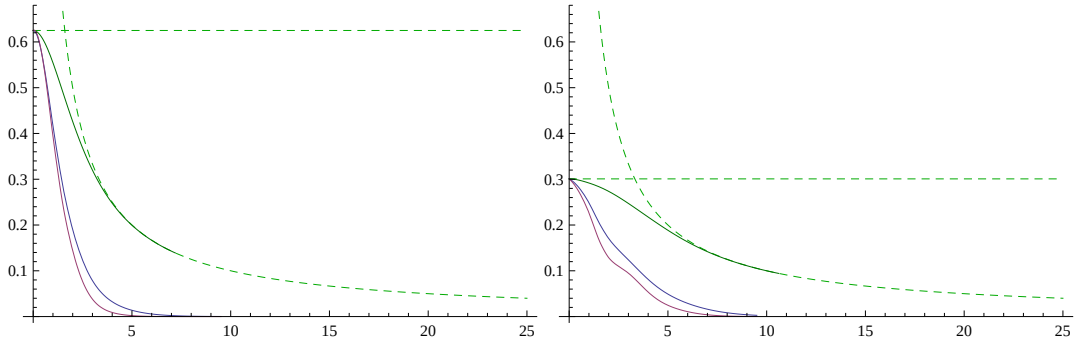


Figure 11.6: This figure shows the interelectron repulsion integral $J_{1s,1s,1s,1s}$ (left) and $J_{2s,2s,2s,2s}$ (right) for a diatomic molecule as a function of the scaled internuclear separation $S = kR$, calculated using the Legendre polynomial expansion method. The atomic value is indicated by a horizontal dotted line. The upper curve corresponds to $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4) = (\mathbf{S}, \mathbf{S}, 0, 0)$. This curve is compared with the asymptotic value, represented by a dashed line, which it approaches for large values of S. The next lower curve represents the case where $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4) = (\mathbf{S}, \mathbf{S}, \mathbf{S}, 0)$, and the lowest curve that where $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4) = (\mathbf{S}, 0, \mathbf{S}, 0)$. These two curves are qualitatively correct (the unevenness of the $J_{2s,2s,2s,2s}$ is physical and due to the nodes), but the accuracy is not satisfactory because of truncation in the Legendre polynomial expansion.

and thus we obtain

$$\begin{aligned} a_{\mu,\mu_1,\mu_2} = {} & n(2k)^{3/2} Y_{l_2,m_2}(\nabla_A) \sum_{q=0}^{l_2} \binom{l_2}{q} (kB)^{l_2-q} (\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^{l_2-q} \\ & \times \sum_{l'} \int d\Omega \, (\hat{\mathbf{x}} \cdot \hat{\mathbf{A}})^q P_{l'}(\hat{\mathbf{x}} \cdot \hat{\mathbf{B}}) Y^*_{l_1,m_1}(\hat{\mathbf{x}}) Y^*_{lm}(\hat{\mathbf{x}}) \\ & \times \sum_{n'=l}^{n} 2^{n'} c^{n'}_{n,l} \sum_{n'_1=l_1}^{n_1} c^{n'_1}_{n_1,l_1} \sum_{n'_2=l_2}^{n_2} c^{n'_2}_{n_2,l_2} \int_0^\infty d(kr) \, (kr)^{n'+n'_1+1} e^{-3kr} v_{l'}[(kr)^{q+n'_2-l_2} e^{-kr}](r, B) \end{aligned} \tag{11.4.35}$$

We have verified that Equation (11.4.35) yields the same results as Equations (11.4.21), (11.4.23), and (11.4.25) respectively, in the special cases to which these equations apply.

11.5 Discussion

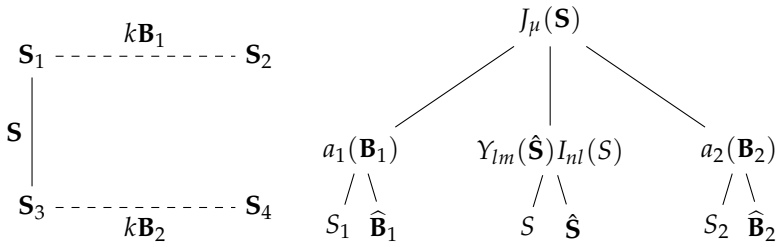
The Legendre-polynomial expansion yields reasonable results, as can be seen in the Figures included here. The curves in the figures that fall rapidly to zero involve two-center densities. The most rapidly falling curve involves two two-center densities, and the least rapidly decreasing curve involves one one-center density and one two-center density. The integrals were generated using Legendre-polynomials P_l up to $l = 4$, and while the results are qualitatively correct, a larger number of terms would be needed to obtain good accuracy.

Looking at the hyperspherical method as a whole, we find that it is extremely efficient in the case of integrals involving only one-center densities (Equation (11.1.1)), where the method is also exact. The reason that the method is so efficient is in part that the integrals

$$I_{nl}(S) = \int_0^\infty dq \, M(q) f_{nl}(q) j_l(qS)$$

in Equation (11.3.9) are evaluated once and for all and stored. Because the angular dependence of J_μ is decoupled from the radial dependence, and is simply a spherical harmonic for each μ, only the radial functions are stored. Since these are one-dimensional functions, we can choose to interpolate the computed closed form expressions using very high numerical precision and store the interpolation functions. This makes evaluation almost instantaneous and eliminates numerical inaccuracy.

The sums in Equations (11.3.6) and (11.3.11) are finite and small when only one-center densities are involved. However, in the case of two-center densities, the coefficients $a_{\mu,\mu_1\mu_2}(\mathbf{B})$ form an infinite series that must be truncated after sufficient accuracy has been obtained. The property that angular and radial dependency can be precalculated and stored separately remains, but we are even more fortunate than that: Also the expansion coefficients $a_{\mu,\mu_1\mu_2}(\mathbf{B})$ are each small sums of radial times angular functions that can be interpolated and stored separately. This, in fact, effectively separates the $4 \times 3 = 12$ coordinates that are the input parameters in the general four-center case:



Thus, despite the intrinsic high dimensionality of the problem, it can be broken down into a sum of radial and angular problems. The angular dependency of $J_\mu(\mathbf{S})$, as well as the three most simple cases of the expansion coefficients $a_{\mu;\mu_1\mu_2}(\mathbf{B})$, are just spherical harmonics, which are so simple that it is most efficient to compute them on the fly. The more complicated angular dependencies of the most difficult type of $a_{\mu;\mu_1\mu_2}$-coefficients in Equation (11.4.35) can be stored either as closed form expressions or as two-dimensional interpolation functions to make evaluation more rapid. This separation property makes the hyperspherical method quite desirable for general three- and four-center computations, and gives us strong motivation to attempt to get better accuracy from the Legendre polynomial expansion.

CHAPTER 12

Evaluation using Gaussian approximations

12.1 Evaluation of many-center interelectron repulsion integrals using harmonic projection

12.1.1 Introductory remarks

The case when two-center densities are involved

$$J_{\tau_1\tau_2;\tau_3,\tau_4} = \int d^3x \int d^3x' \chi^*_{\mu_1}(\mathbf{x}-\mathbf{X}_1)\chi^*_{\mu_2}(\mathbf{x}-\mathbf{X}_2)\frac{1}{|\mathbf{x}-\mathbf{x}'|}\chi^*_{\mu_3}(\mathbf{x}'-\mathbf{X}_3)\chi_{\mu_4}(\mathbf{x}'-\mathbf{X}_4) \tag{12.1.1}$$

poses difficulties. We cannot evaluate the repulsion integrals exactly, and must use expansions. As previously, when we are dealing with Coulomb Sturmians, we can rewrite this equation in terms of the scaled coordinates $\mathbf{s} = k\mathbf{x}$, $\mathbf{S}_i = k\mathbf{X}_i$:

$$J_{\tau_1\tau_2;\tau_3,\tau_4} = k\int d^3s \int d^3s' \chi^*_{\mu_1}(\mathbf{s}-\mathbf{S}_1)\chi^*_{\mu_2}(\mathbf{s}-\mathbf{S}_2)\frac{1}{|\mathbf{s}-\mathbf{s}'|}\chi^*_{\mu_3}(\mathbf{s}'-\mathbf{S}_3)\chi_{\mu_4}(\mathbf{s}'-\mathbf{S}_4) \tag{12.1.2}$$

A method for accurate evaluation of such integrals has been developed, and will be described below.

We will show below that the Coulomb Sturmian atomic orbitals can be expressed in terms of Gaussian expansions, where the coefficients $\gamma_{n,i}$ are universals that need never be changed despite changes in scaling due to changes in the value of k. The expansions used have the form:

$$s^n e^{-s} \approx \sum_i \gamma_{n,i} e^{-a_i s^2} \tag{12.1.3}$$

Figure 12.1 shows the Gaussian expansions of $s^n e^{-s} := e^{-kr}$, while the following table shows the universal Gaussian exponents and coefficients.

The method discussed below also makes use of displaced regular solid harmonics to represent Coulomb Sturmian atomic orbitals located at various points in space. In calculating the Fourier transforms of two-center densities, a closely related differential operator is introduced. It is formed by replacing the coordinate $\mathbf{s}$ in a displaced regular solid harmonic by $i\nabla_q$. Here ∇_q is a gradient with respect to the scaled momentum vector $\mathbf{q}$.

In the method discussed below, the heavy work can be performed off-line, and hence it is possible to use enough Gaussians to accurately represent exponential-type orbitals. This is in contrast

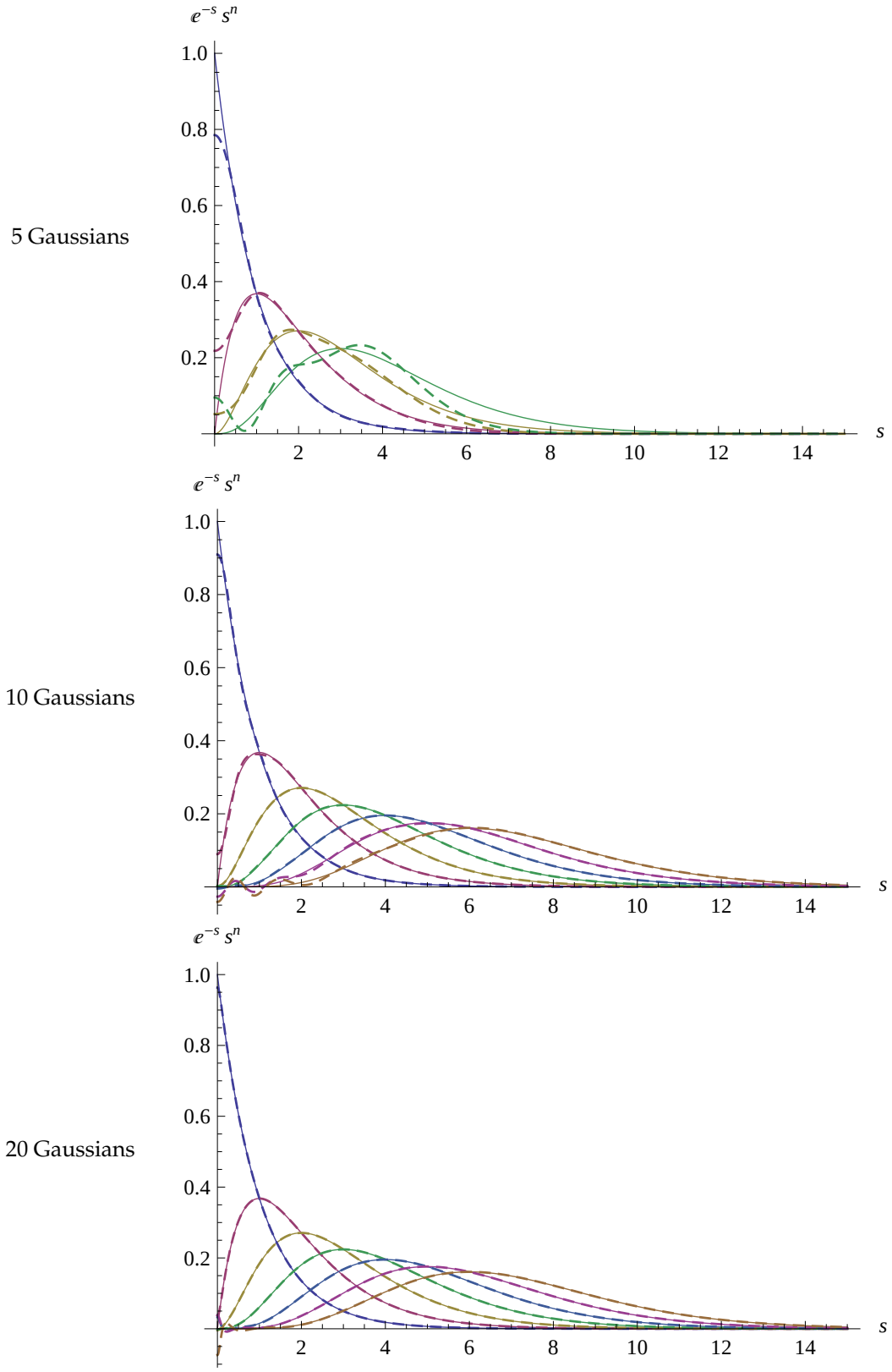


Figure 12.1: This figure shows the Gaussian expansion $s^n e^{-s} \approx \sum_i \gamma_{n,i} e^{-a_i s^2}$ compared to the exact functions for three different levels of approximation. As the power of s grows, it becomes progressively more difficult to represent the exponentials. However, as we increase the number of Gaussian functions, we obtain good approximations for most of the range, but the cusp at $s = 0$ requires many high-exponent Gaussians to approximate.

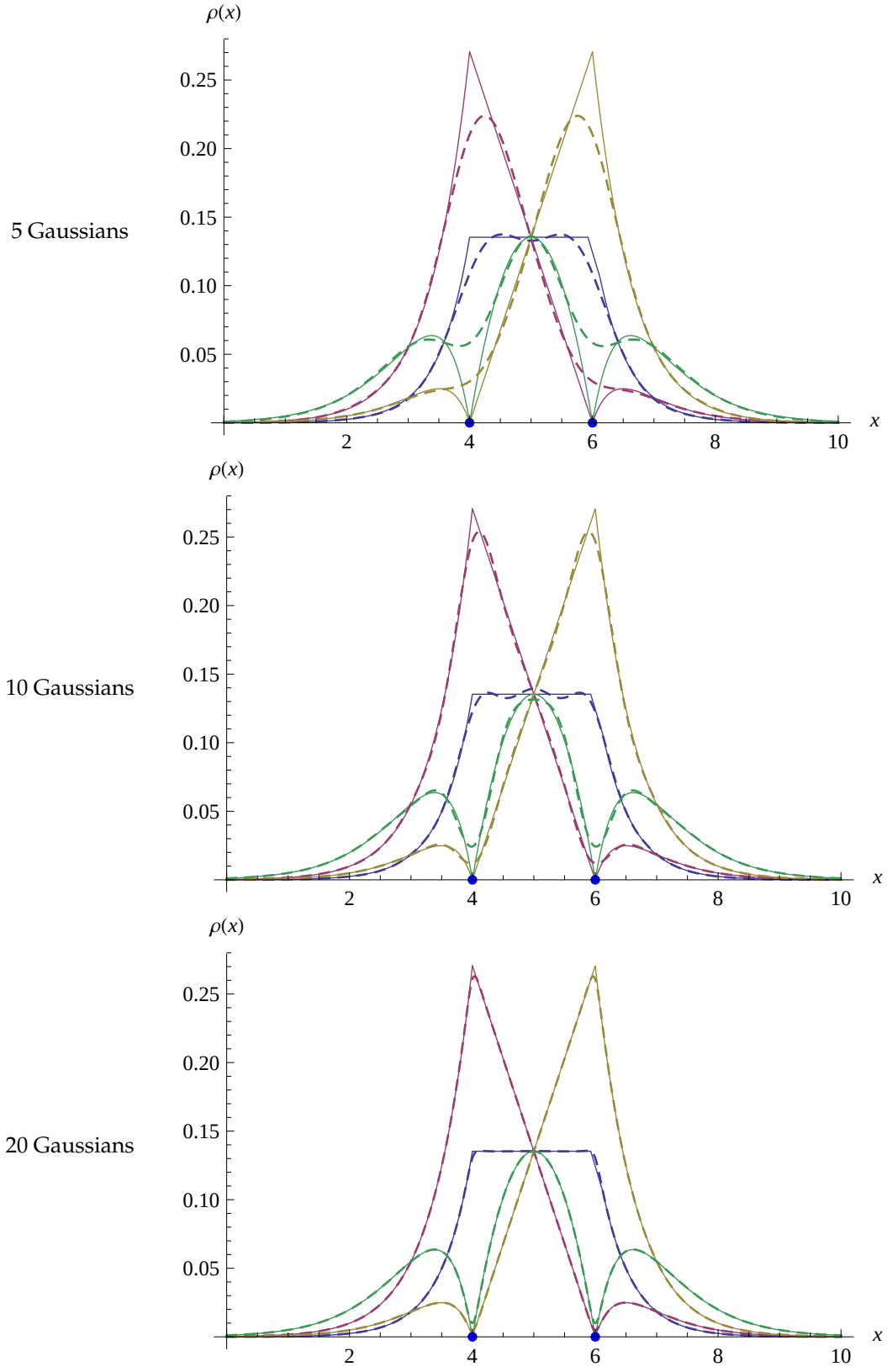


Figure 12.2: This figure shows products of the exponential functions shown in Figure 12.1 centered on $x = 4$ and $x = 6$ respectively. We see that the relative error in the approximate density is more pronounced than for the single orbitals. The sharp features are due to the exponential cusps, which are reproduced poorly in Gaussian expansions. Nevertheless, since calculations are done off-line, we have the luxury of using a large number of Gaussians in the expansion and obtain good results.

a_i	$\gamma_{0,i}$	$\gamma_{1,i}$	$\gamma_{2,i}$
512/100	0.474589	-0.456553	-0.011253
256/100	-0.409842	0.420846	-0.135640
128/100	0.522704	-0.461490	-0.030952
64/100	-0.028869	0.157189	-0.390496
32/100	0.237377	0.008340	-0.284720
16/100	0.074194	0.248277	0.001174
8/100	0.039810	0.147977	0.631545
4/100	-0.001091	0.025882	0.224411
2/100	0.000808	-0.001018	0.000462
1/100	-0.000129	0.000170	0.000468

Table 12.1: Expansion coefficients for expressing $s^n e^{-s}$ in terms of $e^{-a_i s^2}$.

with conventional methods making use of Gaussians, where the integrals are not universal but different for each system, and hence must be carried out on-line. Furthermore, in the method used here, Gaussian expansions are used only to treat two-center densities in interelectron repulsion integrals. No other use is made of Gaussian expansions, and all other integrals are exact. In treating two-center densities, the Gaussian expansion method discussed below has proved to be superior in accuracy to the Legendre polynomial expansion method, at least with the number of terms that have been used until now in the Legendre polynomial expansions.

Expansion of Coulomb Sturmian densities in terms of Gaussians

Let us first consider expansions of Gaussians in terms of Coulomb Sturmian radial functions:

$$e^{-a_i s^2} \approx \sum_n R_{n0}(s) c_{in} \tag{12.1.4}$$

The potential weighted orthogonality relation (10.1.6) for Coulomb Sturmians yields

$$\int_0^\infty ds\, s\, R_{n'0}(s) R_{n0}(s) = \frac{\delta_{n'n}}{n} \tag{12.1.5}$$

Multiplying both sides of Equation (12.1.4) by $s \cdot R_{n'0}(s)$ and integrating over s from 0 to infinity, we then get from Equation (12.1.5)

$$c_{in'} = n' \int_0^\infty ds\, s\, R_{n'0}(s) e^{-a_i s^2} \tag{12.1.6}$$

Having obtained the coefficients c_{jn}, we can obtain the coefficients for the converse expansion simply by inverting the matrix c_{in}:

$$R_{n0}(s) \approx \sum_i c_{ni}^{-1} e^{-a_i s^2} \tag{12.1.7}$$

from which we can easily extract a set of coefficients γ_{ni} such that

$$s^n e^{-s} \approx \sum_i \gamma_{ni} e^{-a_i s^2} \tag{12.1.8}$$

This is because, for any fixed l, the polynomial parts of $\{ R_{nl} \mid n \geq l+1 \leq m \}$ form a basis for $\mathcal{P}_{m-l}$, the set of polynomials of degree no greater than $m-l$. We can write

$$R_{n0}(s) = p(s) e^{-s} = e^{-s} \sum_{j=0}^{n-1} b_{n-1,j} s^j = \sum_i \gamma_{n-1,i} e^{-a_i s^2} \tag{12.1.9}$$

whereby we obtain

$$\begin{aligned} \gamma_{0i} &= \frac{c_{0i}^{-1}}{a_{00}} \\ \gamma_{n,i} &= \frac{1}{a_{nn}} \left(c_{ni}^{-1} - \sum_{m=0}^{n-1} a_{nm} \gamma_{mi} \right) \end{aligned} \tag{12.1.10}$$

To cover both short range and long range behavior, it is useful to let the a_i's grow exponentially. If we choose the Gaussian exponents to be $a_i = \frac{2^{i-1}}{100}$, then we obtain the coefficients given in Table 12.1. As mentioned above, the Coulomb Sturmians are automatically scaled, the scaling factor being k. The expansions also are scaled automatically, and thus the coefficients in the table can be used for all occasions, regardless of whether the wave functions are diffuse or contracted.

Then

$$\begin{aligned} \chi_{1,0,0}(k\mathbf{x} - k\mathbf{X}_a) &= \frac{k^{3/2}}{\sqrt{\pi}} e^{-k|\mathbf{x}-\mathbf{X}_a|} \approx \frac{k^{3/2}}{\sqrt{\pi}} \sum_i \gamma_{0,i} e^{-a_i (k\mathbf{x} - k\mathbf{X}_a)^2} \\ \chi_{2,0,0}(k\mathbf{x} - k\mathbf{X}_a) &= \frac{k^{3/2}}{\sqrt{\pi}} \left(1 - k|\mathbf{x} - \mathbf{X}_a|\right) e^{-k|\mathbf{x}-\mathbf{X}_a|} \\ &\approx \frac{k^{3/2}}{\sqrt{\pi}} \sum_i \left(\gamma_{0,i} - \gamma_{1,i}\right) e^{-a_i (k\mathbf{x} - k\mathbf{X}_a)^2} \\ \chi_{2,1,0}(k\mathbf{x} - k\mathbf{X}_a) &= \frac{k^{3/2}}{\sqrt{\pi}} \left(kz - kZ_a\right) e^{-k|\mathbf{x}-\mathbf{X}_a|} \\ &\approx \frac{k^{3/2}}{\sqrt{\pi}} \left(kz - kZ_a\right) \sum_i \gamma_{0,i} e^{-a_i (k\mathbf{x} - k\mathbf{X}_a)^2} \end{aligned} \tag{12.1.11}$$

and so on. We now introduce the regular solid harmonics R_l^m defined by

$$\mathsf{R}_l^m(\mathbf{x}) := \sqrt{\frac{4\pi}{2l+1}} r^l Y_{lm}(\hat{\mathbf{x}}) \tag{12.1.12}$$

The first few of these are

$$\begin{aligned} \mathsf{R}_0^0(\mathbf{x}) &= 1 \\ \mathsf{R}_1^{-1}(\mathbf{x}) &= (x - iy)/\sqrt{2} \\ \mathsf{R}_1^0(\mathbf{x}) &= z \\ \mathsf{R}_1^1(\mathbf{x}) &= -(x + iy)/\sqrt{2} \end{aligned} \tag{12.1.13}$$

Expressed in terms of the regular solid harmonics, the Coulomb Sturmian basis functions can be written as

$$\begin{aligned} \chi_{n,l,m}(\mathbf{x}) &= R_{n,l}(s) Y_{l,m}(\hat{\mathbf{x}}) \\ &= \sqrt{\frac{2l+1}{4\pi}} R_{n,l}(s) s^{-l} \mathsf{R}_l^m(k\mathbf{x}) \\ &\approx \sum_i \Gamma_{n,l,i} e^{-a_i |k\mathbf{x}|^2} \mathsf{R}_l^m(k\mathbf{x}) \end{aligned} \tag{12.1.14}$$

where the coefficients $\Gamma_{n,l,i}$ are defined by the relationship

$$\sqrt{\frac{2l+1}{4\pi}} R_{n,l}(s) s^{-l} \approx \sum_i \Gamma_{n,l,i} e^{-a_i s^2} \tag{12.1.15}$$

Then

$$\chi_{n,l,m}(\mathbf{x} - \mathbf{X}_a) \approx \sum_i \Gamma_{n,l,i}\, e^{-a_i |k\mathbf{x} - k\mathbf{X}_a|^2} \mathsf{R}_l^m(k\mathbf{x} - k\mathbf{X}_a) \tag{12.1.16}$$

and

$$\begin{aligned} \rho_{\tau,\tau'}(\mathbf{x}) &= \chi^*_{n,l,m}(\mathbf{x} - \mathbf{X}_a) \chi_{n',l',m'}(\mathbf{x} - \mathbf{X}_{a'}) \\ &\approx \sum_i \Gamma_{n,l,i}\, e^{-a_i |k\mathbf{x} - k\mathbf{X}_a|^2} \mathsf{R}_l^m(k\mathbf{x} - k\mathbf{X}_a)^* \\ &\times \sum_j \Gamma_{n',l',j}\, e^{-a_j |k\mathbf{x} - k\mathbf{X}_a|^2} \mathsf{R}_{l'}^{m'}(k\mathbf{x} - k\mathbf{X}_{a'}) \\ &= \mathsf{R}_l^m(k\mathbf{x} - k\mathbf{X}_a)^* \mathsf{R}_{l'}^{m'}(k\mathbf{x} - k\mathbf{X}_{a'}) \\ &\times \sum_i \sum_j \Gamma_{n',l',j} \Gamma_{n,l,i}\, e^{-a_i |k\mathbf{x} - k\mathbf{X}_a|^2} e^{-a_j |k\mathbf{x} - k\mathbf{X}_{a'}|^2} \\ &= \mathsf{R}_l^m(k\mathbf{x} - k\mathbf{X}_a)^* \mathsf{R}_{l'}^{m'}(k\mathbf{x} - k\mathbf{X}_{a'}) \\ &\times \sum_i \sum_j \Gamma_{n',l',j} \Gamma_{n,l,i}\, f_{i,j}(k\mathbf{x}) \\ &= \mathsf{R}_l^m(\mathbf{s} - \mathbf{S}_a)^* \mathsf{R}_{l'}^{m'}(\mathbf{s} - \mathbf{S}_{a'}) \\ &\times \sum_i \sum_j \Gamma_{n',l',j} \Gamma_{n,l,i}\, f_{i,j}(\mathbf{s}) \end{aligned} \tag{12.1.17}$$

where

$$\begin{aligned} f_{i,j}(\mathbf{s}) &= e^{-a_i |\mathbf{s} - \mathbf{S}_a|^2 - a_j |\mathbf{s} - \mathbf{S}_{a'}|^2} \\ &= e^{-\frac{a_i a_j}{a_i + a_j} |\mathbf{S}_a - \mathbf{S}_{a'}|^2} e^{-(a_i + a_j) |\mathbf{s} - \mathbf{S}_c|^2} \end{aligned} \tag{12.1.18}$$

and

$$\mathbf{S}_c := \frac{a_i \mathbf{S}_a + a_j \mathbf{S}_{a'}}{a_i + a_j} \tag{12.1.19}$$

We can now write the Fourier transformed representation of the density $\rho_{\tau,\tau'}$ in terms of the solid harmonics and $f_{i,j}$:

$$\begin{aligned}
\rho^t_{\tau,\tau'}(\mathbf{q}) &= \frac{1}{(2\pi)^{3/2}} \int d^3x\, e^{-i\mathbf{p}\cdot\mathbf{x}}\, \mathsf{R}^m_l(\mathbf{s}-\mathbf{S}_a)^* \mathsf{R}^{m'}_{l'}(\mathbf{s}-\mathbf{S}_{a'}) \\
&\times \sum_i \sum_j \Gamma_{n',l',j} \Gamma_{n,l,i}\, f_{i,j}(\mathbf{s}) \\
&= \frac{1}{(2\pi)^{3/2}} \sum_i \sum_j \tilde{\Gamma}_{n',l',j} \tilde{\Gamma}_{n,l,i} \\
&\times \int d^3s\, e^{-i\mathbf{q}\cdot\mathbf{s}}\, \mathsf{R}^m_l(\mathbf{s}-\mathbf{S}_a)^* \mathsf{R}^{m'}_{l'}(\mathbf{s}-\mathbf{S}_{a'})\, f_{i,j}(\mathbf{s})
\end{aligned} \tag{12.1.20}$$

where

$$\begin{aligned}
\tilde{\Gamma}_{n,l,i} &:= \frac{1}{k^{3/2}} \Gamma_{n,l,i} \\
\mathbf{q} &:= \frac{1}{k} \mathbf{p} \\
d^3s &:= k^3\, d^3x
\end{aligned} \tag{12.1.21}$$

We now notice that

$$\begin{aligned}
f^t_{i,j}(\mathbf{q}) &:= \frac{1}{(2\pi)^{3/2}} \int d^3s\, e^{-i\mathbf{q}\cdot\mathbf{s}}\, f_{i,j}(\mathbf{s}) \\
&= C_{i,j} e^{-\alpha_{i,j} q^2 - i\mathbf{q}\cdot\mathbf{S}_{c,i,j}}
\end{aligned} \tag{12.1.22}$$

with

$$C_{i,j} := \frac{1}{[2(a_i+a_j)]^{3/2}} \exp\left[-\frac{a_i a_j}{a_i + a_j} |\mathbf{S}_a - \mathbf{S}_b|^2 \right] \tag{12.1.23}$$

and

$$\alpha_{i,j} := \frac{1}{4(a_i+a_j)} \qquad \mathbf{S}_{c,i,j} := \frac{a_i \mathbf{S}_a + a_j \mathbf{S}_{a'}}{a_i + a_j} \tag{12.1.24}$$

Thus, if we can construct an operator $\mathsf{R}^m_l(i\nabla_q - \mathbf{S}_a)$ such that

$$\mathsf{R}^m_l(i\nabla_q - \mathbf{S}_a) e^{-i\mathbf{q}\cdot\mathbf{s}} = \mathsf{R}^m_l(\mathbf{s}-\mathbf{S}_a) e^{-i\mathbf{q}\cdot\mathbf{s}} \tag{12.1.25}$$

then we will have

$$\begin{aligned}
&\frac{1}{(2\pi)^{3/2}} \int d^3s\, e^{-i\mathbf{q}\cdot\mathbf{s}}\, \mathsf{R}^m_l(\mathbf{s}-\mathbf{S}_a)^* \mathsf{R}^{m'}_{l'}(\mathbf{s}-\mathbf{S}_{a'})\, f_{i,j}(\mathbf{s}) \\
&= \mathsf{R}^m_l(i\nabla_q - \mathbf{S}_a)^* \mathsf{R}^{m'}_{l'}(i\nabla_q - \mathbf{S}_{a'}) f^t_{i,j}(\mathbf{q}) \\
&:= P^{i,j}_{l,m;l'm'}(\mathbf{q}) f^t_{i,j}(\mathbf{q})
\end{aligned} \tag{12.1.26}$$

We can indeed construct such an operator from the displaced solid harmonic $\mathsf{R}^m_l(\mathbf{s}-\mathbf{S}_a)$ if we replace $\mathbf{s}$ everywhere by $i\nabla_q$. Then, combining (12.1.26) with (12.1.20), we obtain

$$\rho^t_{\tau,\tau'}(\mathbf{q}) = \sum_i \sum_j \tilde{\Gamma}_{n',l',j} \tilde{\Gamma}_{n,l,i} P^{i,j}_{l,m;l'm'}(\mathbf{q}) f^t_{i,j}(\mathbf{q}) \tag{12.1.27}$$

where $P^{i,j}_{l,m;l'm'}(\mathbf{q})$ is a polynomial. We now remember that

$$\begin{aligned} J_{\tau,\tau';\tau'',\tau'''} &= 4\pi \int d^3p \frac{1}{p^2} \rho^t_{\tau,\tau'}(\mathbf{p}) \rho^t_{\tau'',\tau'''}(-\mathbf{p}) \\ &= 4\pi \int_0^\infty dp \int d\Omega_p \, \rho^t_{\tau,\tau'}(\mathbf{p}) \rho^t_{\tau'',\tau'''}(-\mathbf{p}) \\ &= 4\pi k \int_0^\infty dq \int d\Omega_q \, \rho^t_{\tau,\tau'}(\mathbf{q}) \rho^t_{\tau'',\tau'''}(-\mathbf{q}) \end{aligned} \tag{12.1.28}$$

and thus

$$\begin{aligned} J_{\tau,\tau';\tau'',\tau'''} &= 4\pi k \int_0^\infty dq \int d\Omega_q \, \rho^t_{\tau,\tau'}(\mathbf{q}) \rho^t_{\tau'',\tau'''}(-\mathbf{q}) \\ &= 4\pi k \sum_{\mathbf{i}} \tilde{\Gamma}_{n,l,i_1} \tilde{\Gamma}_{n',l',i_2} \tilde{\Gamma}_{n'',l''',i_3} \tilde{\Gamma}_{n''',l''',i_4} \\ &\times \int_0^\infty dq \int d\Omega_q \, P^{i_1,i_2}_{l',m';l,m}(\mathbf{q}) \, P^{i_3,i_4}_{l'',m'';l''',m'''}(-\mathbf{q}) f^t_{i_1,i_2}(\mathbf{q}) f^t_{i_3,i_4}(-\mathbf{q}) \end{aligned} \tag{12.1.29}$$

Then, combining (12.1.29) and (12.1.22), we can write:

$$\begin{aligned} J_{\tau_1,\tau_2;\tau_3,\tau_4} &= 4\pi k \sum_{\mathbf{i}} \tilde{\Gamma}_{n_1,l_1,i_1} \tilde{\Gamma}_{n_2,l_2,i_2} \tilde{\Gamma}_{n_3,l_3,i_3} \tilde{\Gamma}_{n_4,l,i_4} C_{i_1,i_2} C_{i_3,i_4} \\ &\times \int_0^\infty dq \, e^{-\alpha_{\mathbf{i}} q^2} \int d\Omega_q \, P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{S_i}} \end{aligned} \tag{12.1.30}$$

where

$$\begin{aligned} \mathbf{S_i} &:= \mathbf{S}c, i_3, i_4 - \mathbf{S}c, i_1, i_2 \\ P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) &:= P^{i_1,i_2}_{l_1,m_1;l_1,m_2}(\mathbf{q}) \, P^{i_3,i_4}_{l_3,m_3;l_4,m_4}(-\mathbf{q}) \\ \alpha_{\mathbf{i}} &:= \frac{1}{4(a_{i_1} + a_{i_2})} + \frac{1}{4(a_{i_3} + a_{i_4})} \end{aligned} \tag{12.1.31}$$

The integral in (12.1.30) can be rewritten as

$$\begin{aligned} I^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{S_i}) &:= \int_0^\infty dq \, e^{-\alpha_{\mathbf{i}} q^2} \int d\Omega_q \, P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) \, e^{i\mathbf{p}\cdot\mathbf{S_i}} \\ &= 4\pi \sum_l i^l \int_0^\infty dq \, e^{-\alpha_{\mathbf{i}} q^2} j_l(qS_{\mathbf{i}}) \sum_{m=-l}^{l} Y_{l,m}(\hat{\mathbf{S}}_{\mathbf{i}}) \int d\Omega_q \, Y^*_{l,m}(\hat{\mathbf{q}}) P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) \end{aligned} \tag{12.1.32}$$

In the special case where $\hat{\mathbf{S}}_{\mathbf{i}} = (0,0,1)$, equation (12.1.32) becomes

$$\begin{aligned} I^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{S_i}) &= 4\pi \sum_l \int_0^\infty dq \, e^{-\alpha_{\mathbf{i}} q^2} j_l(qS_{\mathbf{i}}) \sqrt{\frac{2l+1}{4\pi}} \int d\Omega_q \, Y_{l,0}(\hat{\mathbf{q}}) P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) \\ &= \sum_l i^l (2l+1) \int_0^\infty dq \, e^{-\alpha_{\mathbf{i}} q^2} j_l(qS_{\mathbf{i}}) \int d\Omega_q \, P_l(q_3/q) P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) \end{aligned} \tag{12.1.33}$$

where P_l is a Legendre polynomial. The angular integral

$$\int d\Omega_q \, P_l(q_3/q) P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) = \sum_t q^t c^{\mathbf{i},t}_{\mathbf{l},\mathbf{m}} \tag{12.1.34}$$

can be evaluated with the help of the harmonic projection operator h_0, and the remaining radial integral over $\mathbf{p}$ has the form:

$$I_{t,l} := \int_0^\infty dq \, q^t \, e^{-\alpha_{\mathbf{i}} q^2} j_l(qS_{\mathbf{i}}) \tag{12.1.35}$$

Integrals of this type can readily be evaluated by Mathematica. Thus we finally obtain

$$J_{\tau_1,\tau_2;\tau_3,\tau_4} = 4\pi k \sum_{\mathbf{i}} \tilde{\Gamma}_{n_1,l_1,i_1} \tilde{\Gamma}_{n_2,l_2,i_2} \tilde{\Gamma}_{n_3,l_3,i_3} \tilde{\Gamma}_{n_4,l_4,i_4} C_{i_1,i_2} C_{i_3,i_4} I^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{R_i}) \tag{12.1.36}$$

where, for the special case where $\hat{\mathbf{S}}_\mathbf{i} = (0,0,1)$,

$$I^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{S_i}) = \sum_l i^l (2l+1) \sum_t I_{t,l}(\alpha_\mathbf{i}, S_\mathbf{i}) c^{\mathbf{i},t}_{\mathbf{l},\mathbf{m}} \tag{12.1.37}$$

while for general directions of $\hat{\mathbf{S}}_\mathbf{i}$,

$$I^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{S_i}) = 4\pi \sum_l \sum_{m=-l}^{l} Y_{l,m}(\hat{\mathbf{S}}_\mathbf{i}) i^l \int_0^\infty dq\, e^{-\alpha_\mathbf{i} q^2} j_l(qS_\mathbf{i}) \int d\Omega_q\, Y^*_{l,m}(\hat{\mathbf{q}}) P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) \tag{12.1.38}$$

For the particular case where $\hat{\mathbf{S}}_\mathbf{i} = (0,0,1)$ and $\mathbf{l} = (0,0,0,0)$, we have

$$I^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{S_i}) = 4\pi\, I_{0,0}(\alpha_\mathbf{i}, S_\mathbf{i}) = \frac{2\pi^2}{S_\mathbf{i}} \mathrm{Erf}\left[\frac{S_\mathbf{i}}{2\sqrt{\alpha_\mathbf{i}}}\right] \tag{12.1.39}$$

with

$$\begin{aligned}
\frac{S_\mathbf{i}}{2\sqrt{\alpha_\mathbf{i}}} &= \frac{|\mathbf{S}c, i_3, i_4 - \mathbf{S}c, i_1, i_2|}{2\sqrt{\alpha_\mathbf{i}}} \\
&= \sqrt{\frac{(a_{i_1}+a_{i_2})(a_{i_3}+a_{i_4})}{a_{i_1}+a_{i_2}+a_{i_3}+a_{i_4}}}\, |\mathbf{S}c, i_3, i_4 - \mathbf{S}c, i_1, i_2| \\
&= \sqrt{\frac{(a_{i_1}+a_{i_2})(a_{i_3}+a_{i_4})}{a_{i_1}+a_{i_2}+a_{i_3}+a_{i_4}}} \left| \frac{a_{i_3}S_3 + a_{i_4}S_4}{a_{i_3}+a_{i_4}} - \frac{a_{i_1}S_1 + a_{i_2}S_2}{a_{i_1}+a_{i_2}} \right|
\end{aligned} \tag{12.1.40}$$

and

$$\begin{aligned}
S_\mathbf{i} &= \left| \frac{a_{i_3}S_3 + a_{i_4}S_4}{a_{i_3}+a_{i_4}} - \frac{a_{i_1}S_1 + a_{i_2}S_2}{a_{i_1}+a_{i_2}} \right| \\
\alpha_\mathbf{i} &= \frac{1}{4} \frac{a_{i_1}+a_{i_2}+a_{i_3}+a_{i_4}}{(a_{i_1}+a_{i_2})(a_{i_3}+a_{i_4})}
\end{aligned} \tag{12.1.41}$$

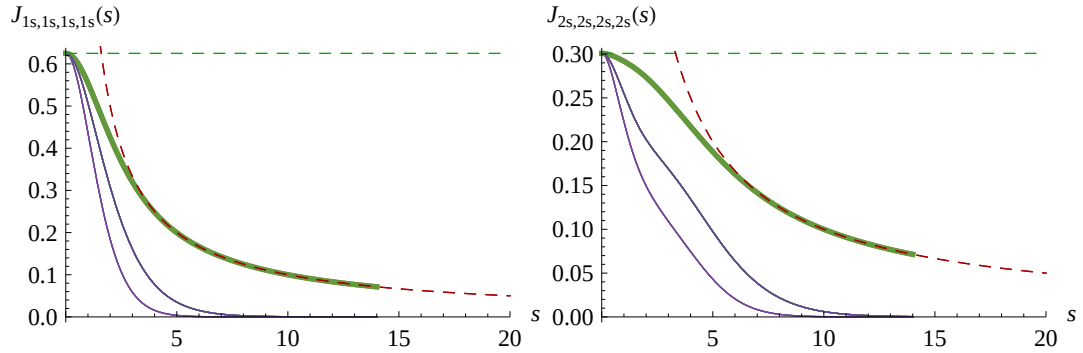


Figure 12.3: This figure, analogous to Figure 11.6, shows the interelectron repulsion integral $J_{1s,1s,1s,1s}$ for a diatomic molecule as a function of the scaled internuclear separation $S = kR$, calculated using the method described in this chapter. The atomic value is indicated by a horizontal dotted line. The upper curve corresponds to $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4) = (\mathbf{S}, \mathbf{S}, 0, 0)$. This curve is compared with the asymptotic value, represented by a red dashed line, which it approaches for large values of S. The next lower curve represents the case where $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4) = (\mathbf{S}, \mathbf{S}, \mathbf{S}, 0)$, and the lowest curve that where $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4) = (\mathbf{S}, 0, \mathbf{S}, 0)$. Due to symmetry, these are the only distinct integrals. For the pure-density integrals $(\mathbf{S}, \mathbf{S}, 0, 0)$, we can compare to the exact function, which shows 4-5 digit accuracy for both integral functions.

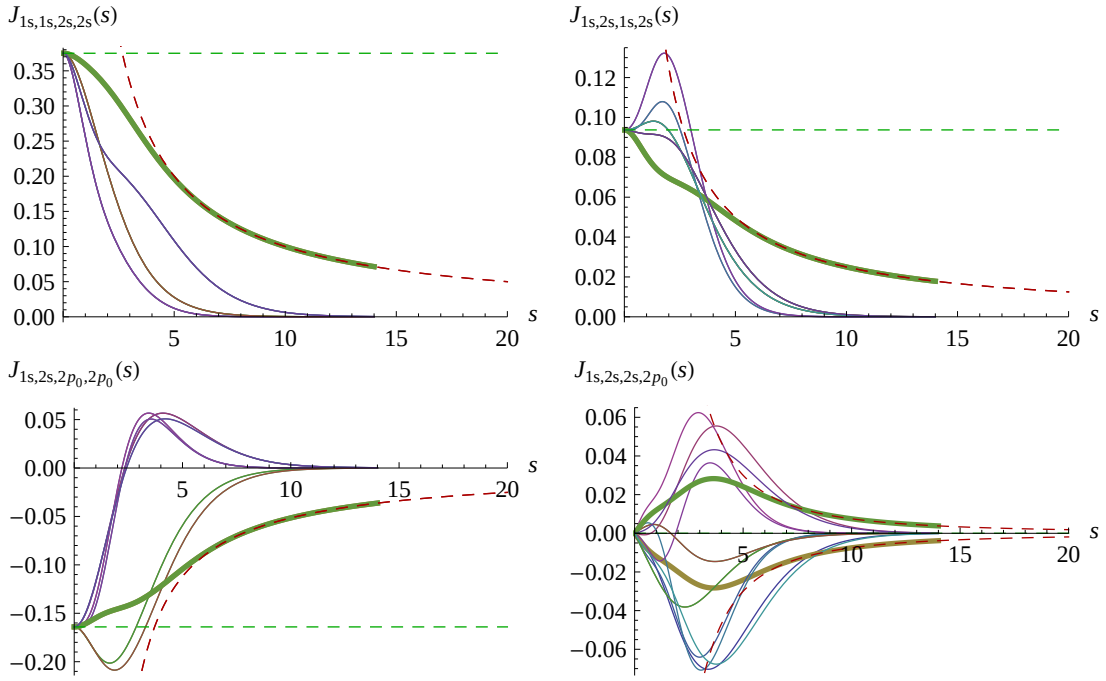


Figure 12.4: More examples of two-center repulsion integrals. In each case, all 16 possible constellations of $(\mathbf{S}_1, \ldots, \mathbf{S}_4)$ are shown. We can see the number of symmetry operations decrease for the more complex cases leading to a greater number of distinct curves. The (long range) pure-density integrals are again shown with fat lines, and the corresponding asymptotic expressions with red, dashed lines. We note that the asymptotic behavior – as the functions approach $S = 0$ and the long range behavior – is correct for all the functions. A separate check shows good agreement in the pure-density case, for which we have exact results.

CHAPTER 13

Software

13.1 Introduction

The supporting software mainly consist of a number of Mathematica modules that implement the supporting mathematical methods described in Chapter 10, and the central equations in Chapters 11 and 12. We have implemented a fully automatic generation of closed form expressions both for the exact multi-center interelectron repulsion integrals, when possible, and for the various levels of approximation. The central building blocks are also generated in closed form and stored. A less complete module for solving molecular problems using isoenergetic basis sets built from Coulomb Sturmians, the method of which is described in Part III, is also included.

The source code can be downloaded from

```
git@sturmian.beanstalkapp.com:/integrals.git
```

in addition, a C++-library is under development, the calculational part of which is generated automatically from the pre-computed closed form expressions for the integrals. Preliminary testing has shown extremely promising efficiency. This library is under heavy development, but for the interested reader, it can be downloaded from

```
git@sturmian.beanstalkapp.com:/libintegrals.git
```

13.2 Computing closed forms for the electron repulsion integrals

Carrying through all computations symbolically, computing exact closed form expressions and simplifying these expressions, is a computationally heavy job. It is thus desirable both to reduce problem sizes and to exploit the parallel nature of the problems.

The number of integrals is automatically reduced using the applicable symmetry relations, and the potentially unique integrals are enumerated, permuted randomly and split between processing nodes. Permutation of the input is required for load balancing, since computation time of the individual integrals varies wildly. Due to the embarrassingly parallel nature of the problem, the integrals may then be evaluated in a completely distributed way. No contact between processing nodes is necessary during computation, and each process writes to a disjoint set of files. Only the final result must reside on a single file system when gathering, postprocessing and storing the integrals.

The integrals used in this thesis were computed on Danmarks Tekniske Universitets (DTU) DCSC Sun UltraSparc machines, requiring about a day's computation time of 20 nodes to generate the entire set, including automatic simplification of the resulting expressions. While this constitutes a great deal of computation time, precalculating, postprocessing and storing integrals is a one-time cost, and the resulting expressions can be evaluated extremely rapidly.

While the various generated integrals may be reduced in numbers by general symmetry arguments, they still exhibit a great deal of redundancy: Not only do we obtain sparsity from selection rules, but the non-zero integrals, especially in the diatomic case, are duplicated to a high degree. To exploit this redundancy, a special data structure based on perfect hashes has been developed and is described in Section 13.5. The improvement over a general sparse array structure is around a factor 5-10.

13.2.1 `Exact`

This module contains code that generates closed form expressions for the one- or two-center integrals that can be computed exactly.

The most important programs are:

generate/generate-atomic.m Generates exact closed form expressions for repulsion integrals in the atomic limit, as well as separate angular and radial integrals.

generate/generate-asymptotes.m Generates exact closed form expressions for the asymptotic limits of all repulsion integrals using the multipole expansion.

generate/generate-00ss-asymp.m Generates exact closed form expressions for two-center integrals of the form

$$\int d^3x_1 \int d^3x_2 \, \rho_1(\mathbf{x}_1 - \mathbf{X}_1) \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|} \rho_2(\mathbf{x}_2 - \mathbf{X}_2)$$

using the methods described in Chapter 11. A convergence radius r_c where the exact integral approaches the asymptotic expression is computed, and a triple (r_c, I_a, I) is stored, where I_a is the asymptotic expression, and I is the exact integral. Outside r_c, the simple asymptotic expression is used.

For diatomics where there is no angular dependence, an interpolation function is generated from the computed closed form expression, evaluated at high numerical accuracy.

13.2.2 `FourierHyper`

This module computes general multi-center, mixed density electron repulsion integrals using the method described in Chapter 11 with Legendre polynomial expansions.

The most important files are the following:

generate-vltable.m Generates the l-terms v_l of the displaced potential $V(r)$, as defined in Equations (10.7.1) and (10.7.4).

generate/generate-diatomic-as.m Generates the expansion coefficients defined in Equations (11.4.21), (11.4.23), (11.4.25), and (11.4.35) as functions of the scaled nuclear distance S.

generate/generate-I74.m Generates the integrals in Equation (11.4.10):

$$\int_0^\infty dr \, e^{-3r} \, r^{n'+1} \, v_l[r^n e^{-kr}](r, B)$$

generate/generate-I94.m Generates the integrals $I_{nl}(S)$ of Equation (11.3.9).

generate/generate-tabcs.m Generates the coefficients for expanding products of hyperspherical harmonics in a sum of single hyperspherical harmonics

$$c_{\mu;\mu_1,\mu_1} = \int d\Omega_4 Y_\mu(\mathbf{u}) Y^*_{\mu_1}(\mathbf{u}) Y_{\mu_2}(\mathbf{u})$$

by way of harmonic projection.

transferas.m Computes closed form expressions for the transfer coefficients $a_{\mu;\mu_1,\mu_2}$ for the four progressive levels of complexity described in Equations (11.4.21), (11.4.23), (11.4.25), and (11.4.35).

generate/generate-Jtable.m Computes all interelection integrals up to the $(n_{\max}, l_{\max}, m_{\max})$ given in `config.m`. The integrals are computed in a loosely coupled distributed fashion, as is described in the beginning of this section.

13.2.3 `SolidGaussian`

This module computes general multi-center, mixed density electron repulsion integrals using the method described in Chapter 12.

The most important files are the following:

solid-gaussian.m Implements a good deal of the plumbing described in Chapter 12.

generate/generate-I31.m Computes the coefficients in Equation (12.1.34) by way of harmonic projection.

generate/generate-I32.m Generates a table of the integrals $I_{tl}(\mathbf{S_i})$ from Equation (12.1.35).

generate/generate-Ppol.m Generates simplified expressions for the momentum-space angular polynomials $P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q})$, defined in Equations (12.1.31) and (12.1.26).

generate/generate-Iilm.m Generates simplified expressions for the integrals $I^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{S_i})$ defined in Equation (12.1.32).

generate/generate-Jtable.m Generates all interelectron integrals up to the $(n_{\max}, l_{\max}, m_{\max})$ given in `config.m`. The integrals are computed in a loosely coupled distributed fashion, as is described in the beginning of this section.

13.2.4 Common methods

harmonics.m Implements the harmonic projection operators in three and four dimensions, using the method based on Theorem 10.3.1.

Plint.m Rapid, closed form angular integration of various types of products consisting of Legendre polynomials, spherical harmonics and unit coordinate polynomials. The method described in Section 10.3.8 is used to automatically produce simple expression.

spherical.m and spherical4.m Implements Cartesian spherical and hyperspherical harmonics.

wulf6.m Contains methods for computing Shibuya-Wulfman integrals, displaced Coulomb Sturmian overlap integrals, Coulomb Sturmian functions, and the gradient spherical harmonics $Y_{lm}(\nabla)$ differential operators.

split.m Distributes work between computational processes.

13.3 `libintegrals` - an efficient interelctron repulsion `C++` library

Programs written in Mathematica provide good examples and intuition about the accuracy and feasibility of our methods. However, in order to compete with state of the art methods, we need efficient implementations. Indeed, the reason for developing these methods is to provide a practical alternative to standard Gaussian methods that should be both faster and more accurate. We are in the process of developing a `C++` library that provides extremely rapid evaluation of both the interelectron repulsion integrals and the integrals that we will need in Part III of the thesis. The Mathematica programs outlined above are used to automatically generate the integrals, which are then processed as described in the remainder of this section. This software is still very much a work in progress and not quite ready for public consumption, but a preliminary version can be cloned from

```
git@sturmian.beanstalkapp.com:/libintegrals.git
```

Integral evaluation time has so far been found to range from tens of clock cycles for the simplest types to hundreds for the more complex types. This is thousands of times faster than the pure Mathematica version, and rapid by any standard.

13.3.1 Diatomics

For diatomics, one can arbitrarily choose the z-axis to coincide with the separation axis. Hereby all integrals become pure functions of the scaled internuclear separation S. Stored interpolation functions can then be used rather than symbolic expressions, and integral evaluation be performed at run time in tens of clock cycles. The interpolation functions are automatically generated; Data points up to the asymptotic convergence radius r_c are evaluated with high numerical working precision using Mathematica. These are then interpolated and stored, as well as loaded and evaluated in the `C++`-library, using the GNU Scientific Library (Galassi).

13.3.2 General two-center integrals

For multi-atomic molecules, two center integrals are no longer pure functions of the nuclear separation. Hence, there is no longer a single axis of nuclear separation along which we can choose the z-axis, and the two-center electron repulsion integrals thus depend not only on the internuclear distance, but also on the two polar angles, θ and ϕ, as is illustrated in Figure 13.1.

Fortunately, in the method based on hyperspherical harmonics, described in Chapter 11, the radial and the angular parts separate cleanly. In the two-center, single-center density case, Equation

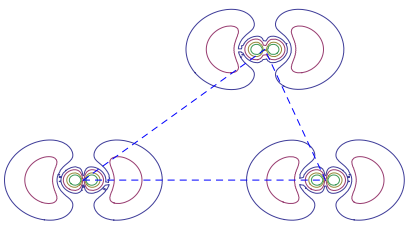


Figure 13.1: This figure illustrates the need to include angles even in two-center integrals for multi-atomic molecules. No two products of orbital pairs have the same dependence on the separation distance.

(11.3.9) states

$$J_\mu = Y_{l,m}(\hat{\mathbf{S}})\ (4\pi)^2\, k \int_0^\infty dq\ M(q) f_{n,l}(q) j_l(qS)$$

The angular and radial parts to J_μ are each very few in number, and the radial part can easily be precalculated and stored as detailed interpolation functions. The spherical harmonics can be initialized at startup for a particular molecular geometry and are then just a list of numbers. In addition, $J_{\mu'\mu}$ of Equation (11.3.11) is the same for all the integrals belonging to a particular pair of atoms, and thus the evaluation of interpolation functions and the two dot products in (11.3.9) and (11.3.11) need only be performed once per pair. In the exact case, we are thus still able to use the precomputed results to get extremely rapid evaluation of electron repulsion integrals.

For the general two-center mixed-density integrals, where approximation is necessary, neither method offers any simplification compared to the three- and four-center cases. They are thus discussed together below.

13.4 Multicenter integrals: Autogenerating efficient C-code

In the cases of three- and four-center interelectron repulsion integrals, the integrals are functions of many variables, in the worst case of four coordinate triplets. Interpolating and storing any non-trivial number of 6-12 dimensional functions is generally infeasible.

Nevertheless, we are not completely out of luck. Even in the mixed-density case, it is possible to separate out the angular parts in the hyperspherical method. The three out of four cases of the Legendre polynomial expansion, (11.4.21), (11.4.23), and (11.4.25) are small sums with only a handful of terms, each a product of a radial part, which can be precomputed and stored as an interpolation function, and a spherical harmonic, which is to be computed on the spot. Even the fourth and most general case in Equation (11.4.35) is a small sum of angular functions that can be precomputed and simplified, multiplied by a pure radial part that can be precomputed and interpolated. Since this gives the expansion coefficients for each two-center density separately, this works the same whether it is a two-, three- or four-center problem, because it effectively separates not only the angular and radial coordinates, but also $\mathbf{x}$ and $\mathbf{x}'$. We can thus employ a very similar scheme to the one used for the exact case above to precompute compact forms that are evaluated efficiently at runtime.

In the case of the Gaussian method described in Chapter 12, we unfortunately are *not* able to separate the terms of Equation (12.1.30)

$$J_{\tau_1,\tau_2;\tau_3,\tau_4} = 4\pi k \sum_{\mathbf{i}} \tilde{\Gamma}_{n_1,l_1,i_1} \tilde{\Gamma}_{n_2,l_2,i_2} \tilde{\Gamma}_{n_3,l_3,i_3} \tilde{\Gamma}_{n_4,l,i_4} C_{i_1,i_2} C_{i_3,i_4}$$
$$\times \int_0^\infty dq\ e^{-\alpha_{\mathbf{i}} q^2} \int d\Omega_q\ P^{\mathbf{i}}_{\mathbf{l},\mathbf{m}}(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{s}_{\mathbf{i}}}$$

into angular and radial factors; Indeed, even if we could, it is not a sum over few terms, but rather a heavy, four-way sum that we very much wish to avoid evaluating at runtime. Thus, when using this method of approximation, separating angular and radial parts is not feasible. Since we have found the Gaussian method to be the most effective and flexible for mixed-center densities in the diatomic case, we do not wish to give up on the method in the general case.

Luckily, there is no cause for despair, albeit some complication does result from this. While we cannot separate variables, and interpolation is infeasible due to the high dimensionality, we are in fact able to compute closed form expressions, parameterized by the atomic positions. Surprisingly, although the expressions result from the large sum, we are still able to simplify them down to a manageable size. The closed form expressions can then simply be evaluated on the fly!

We are developing a small library for Mathematica to perform automatic program transformation and C-code generation. It is not a general purpose code generation library, but has been written

specially to generate very fast C-code specifically from the Mathematica expressions that arise for the various types of integrals with which we have been working, with a set of specialized term rewriting functions for each type of integral that exploits the common structure of that particular type. This has been applied both to the Shibuya-Wulfman integrals, displaced overlap integrals and to the two-center (but angular dependent) interelectron repulsion integrals. The three- and four-center cases are yet to be done, but results from the C-code generated from the preliminary tests have been incredibly promising: We are able to do quad-precision evaluation with computation times of individual integrals, with angular momentum up to $l = 4$, in the range of 200-800 clock cycles.

13.5 Storing and accessing heavily redundant data: Perfect hashes

As mentioned above, most of the integral matrices that we wish to precompute and store exhibit a great deal of redundancy. While some of this redundancy is due to sparsity, the bulk of it in fact stems from repeated non-zero terms. A standard sparse matrix type will thus not yield a sufficiently compact representation. Table 13.1 gives an example of the numbers involved. The table shows the total number of atomic angular integrals, the number of nonzero values, and the number of distinct values. The difference is dramatic!

$l_{\max}$	0	1	2	3	4	5
Number M of (l,m)-pairs	1	4	9	16	25	36
M^4	1	256	6561	65536	390625	1679616
Number of nonzero values	1	39	559	3972	18857	66954
Number of unique values	1	10	79	441	1862	6376

Table 13.1: Sparsity and redundancy of the angular integrals entering into the interelectron repulsion integrals. While the reduction in big-$\mathcal{O}$ growth for the number of nonzero values and unique values does not look overly impressive on paper (both are $\Theta(l_{\max}^7)$, compared to $\Theta(l_{\max}^8)$ for M^4), the actual savings are sizeable.

A second important feature of the precomputed integral matrices is that there is only a single operation on them that we require to be efficient, and that is random access reads. Once generated, these matrices are never updated, and in general, reads from them are not sequential. Thus, a row or column major representation does nothing to optimize access patterns. However, reads are frequent, and the matrices are large, so look ups must be rapid. Additionally, the individual values are quite memory consuming, for example complex expressions or interpolation functions.

In summary, we need to represent a small set of distinct values that must be accessed from a huge index space – the interelectron repulsion matrices given M basis functions are $M \times M \times M \times M$ – in which many indices correspond to the same value. The structure should be as small as possible, random access reads must be as fast as possible, and writes are never performed, i.e. the structures are static. This is a perfect match for a perfect hash. Near-minimal perfect hashes that have constant access times can be generated automatically. We use Bob Jenkins' minimal perfect hash generator `perfect` (Jenkins [1996–2006]), which is able to generate minimal perfect hashes for 32-bit or 64-bit words. Index tuples corresponding to non-zero terms are marshalled into 32 or 64 bit words, and `perfect` is run to generate a corresponding minimal or near-minimal perfect hash that maps marshalled keys to indices into the set of unique data values.

At run-time, two operations are necessary for lookup: Marshalling the key, which is generally done using a few bit-wise operations, and the hash lookup:

$$\text{key} \xrightarrow{\text{marshal}} \text{integer} \xrightarrow{\text{hash}} \text{index} \xrightarrow{\text{lookup}} \text{value} \tag{13.5.1}$$

There is thus very little extra computational overhead compared to flat, random access arrays: As can be seen in Listing 13.1, retrieving an index into the value table requires only 12 machine operations on registers and two memory loads to perform the hash, then one additional load to retrieve the index from the index table. In total, even with huge sparse matrices, lookup time is in tens of clock cycles.

Listing 13.1: Hashing function for 32-bit integer keys produced by Bob Jenkins' perfect hash generator on 50000 distinct keys and index space size 2^{32}.

```
uint32_t phash(uint32_t val)
{
  ub4 a, b, rsl;
  val += 0xefe9566e;
  val ^= (val >> 16);
  val += (val << 8);
  val ^= (val >> 4);
  b = val & 0x7fff;
  a = (val + (val << 17)) >> 17;
  rsl = (a^scramble[tab[b]]);
  return rsl;
}
```

Listing 13.2: Interface to the `PerfectMap` class. All data is generated before compile time.

```
template class PerfectMap<typename key_t, typename value_t> {
  public:

  virtual uint32_t marshal(const key_t&) = 0;

  const value_t& operator[](const key_t& key){
    return this->operator[](marshal(key));
  }

  const value_t& operator[](uint32_t k){
    uint32_t hash = phash(k);
    uint32_t index = indextable[hash];
    return valuetable[index];
  }
  ⋮

  private:
    uint32_t *indextable;
    value_t *valuetable;

    uint32_t phash();
    uint16_t *scramble;
    uint8_t *tab;
};
```

Part III

Molecular Sturmians: First steps

CHAPTER 14

Molecular orbitals based on Coulomb Sturmians

14.1 The one-electron secular equation

Molecular orbitals may be represented as superpositions of Coulomb Sturmian basis functions centered on the nuclei of a molecule. These basis functions are an example of Exponential-Type Orbitals (ETO's), and calculations using them can potentially be much more accurate than calculations based on Gaussians. In this approach to molecular orbital theory, we search for solutions to the one-electron Schrödinger equation

$$\left[-\frac{1}{2}\nabla^2 + v(\mathbf{x}) - \epsilon_\zeta\right] \varphi_\zeta(\mathbf{x}) = 0 \tag{14.1.1}$$

where $v(\mathbf{x})$ is the Coulomb attraction potential of the nuclei:

$$v(\mathbf{x}) = -\sum_a \frac{Z_a}{|\mathbf{x} - \mathbf{X}_a|} \tag{14.1.2}$$

Let us represent the molecular orbitals $\varphi_\zeta(\mathbf{x})$ as superpositions of Coulomb Sturmian atomic orbitals centered on the various atoms of the molecule. To do this it is convenient to introduce a notation where τ stands for a set of four indices, the first three being the quantum numbers of a one-electron Coulomb Sturmian basis function of the type discussed in Part II, Section 10.1, while the final index, a, is the index of the nucleus on which the atomic orbital is centered:

$$\tau \equiv (n, l, m, a) \tag{14.1.3}$$

In this notation we can write

$$\chi_\tau(\mathbf{x}) \equiv \chi_{nlm}(\mathbf{x} - \mathbf{X}_a) \tag{14.1.4}$$

A molecular orbital is then represented by a superposition of the form

$$\varphi_\zeta(\mathbf{x}) = \sum_\tau \chi_\tau(\mathbf{x}) C_{\tau,\zeta} \tag{14.1.5}$$

Coulomb Sturmian basis functions

$$\chi_{nlm}(\mathbf{x}) = R_{n,l}(r) Y_{lm}(\theta, \phi) \tag{14.1.6}$$

have exactly the same form as the familiar hydrogenlike atomic orbitals, except in the radial part, $R_{n,l}(r)$, the factor Z/n is replaced by a constant, k. The first few Coulomb Sturmian radial functions are

$$\begin{aligned} R_{1,0}(r) &= 2k^{3/2}e^{-kr} \\ R_{2,0}(r) &= 2k^{3/2}(1-kr)e^{-kr} \\ R_{2,1}(r) &= \frac{2k^{3/2}}{\sqrt{3}}\, kr\, e^{-kr} \end{aligned} \tag{14.1.7}$$

The reader can verify that these are precisely the same as hydrogenlike atomic orbitals with the replacement $Z/n \to k$. We now substitute the superposition (14.1.5) into the one-electron Schrödinger equation (14.1.1)

$$\sum_{\tau} \left[-\frac{1}{2}\nabla^2 + \frac{1}{2}k^2 + v(\mathbf{x}) \right] \chi_{\tau}(\mathbf{x}) C_{\tau,\zeta} = 0 \tag{14.1.8}$$

with

$$\epsilon_{\zeta} \equiv -\frac{1}{2}k^2 \tag{14.1.9}$$

Next, we make use of the fact that each of the Coulomb Sturmian atomic orbitals obeys a one-electron Schrödinger equation of the form

$$\left[-\frac{1}{2}\nabla^2 + \frac{1}{2}k^2 - \frac{nk}{|\mathbf{x}-\mathbf{X}_a|} \right] \chi_{nlm}(\mathbf{x}-\mathbf{X}_a) = 0 \tag{14.1.10}$$

Using (14.1.8) to replace $-\frac{1}{2}\nabla^2 + \frac{1}{2}k^2$ by $-v(x)$ in Equation (14.1.10), and taking the inner product with a conjugate function in our basis set, we obtain

$$\sum_{\tau} \int d^3x\, \chi^*_{\tau'}(\mathbf{x}) \left[v(\mathbf{x}) + \frac{nk}{|\mathbf{x}-\mathbf{X}_a|} \right] \chi_{\tau}(\mathbf{x}) C_{\tau,\zeta} = 0 \tag{14.1.11}$$

With the notation

$$\mathfrak{W}_{\tau',\tau} \equiv -\frac{1}{k} \int d^3x\, \chi^*_{\tau'}(\mathbf{x}) v(\mathbf{x}) \chi_{\tau}(\mathbf{x}) \tag{14.1.12}$$

and

$$\mathfrak{S}_{\tau',\tau} \equiv \frac{n}{k} \int d^3x\, \chi^*_{\tau'}(\mathbf{x}) \frac{1}{|\mathbf{x}-\mathbf{X}_a|} \chi_{\tau}(\mathbf{x}) \tag{14.1.13}$$

we obtain a secular equation of the form

$$\sum_{\tau} \left[\mathfrak{W}_{\tau',\tau} - k\mathfrak{S}_{\tau',\tau} \right] C_{\tau,\zeta} = 0 \tag{14.1.14}$$

The integrals $\mathfrak{S}_{\tau',\tau}$ are the Shibuya-Wulfman integrals discussed in Part II, but here written in a slightly form. Methods for their evaluation are discussed in Part II, Section 10.5. It can be shown (Koga and Matsuhashi [1988]) that the matrix elements of the many-center potential $\mathfrak{W}_{\tau',\tau}$ can be expressed in terms of the Shibuya-Wulfman integrals by means of the sum rule

$$\mathfrak{W}_{\tau',\tau} = \sqrt{\frac{n'n}{Z_{a'}Z_a}} \sum_{\tau''} K_{\tau',\tau''} K_{\tau'',\tau} \tag{14.1.15}$$

where

$$K_{\tau',\tau} \equiv \sqrt{\frac{Z_{a'}Z_a}{n'n}} \mathfrak{S}_{\tau',\tau} \tag{14.1.16}$$

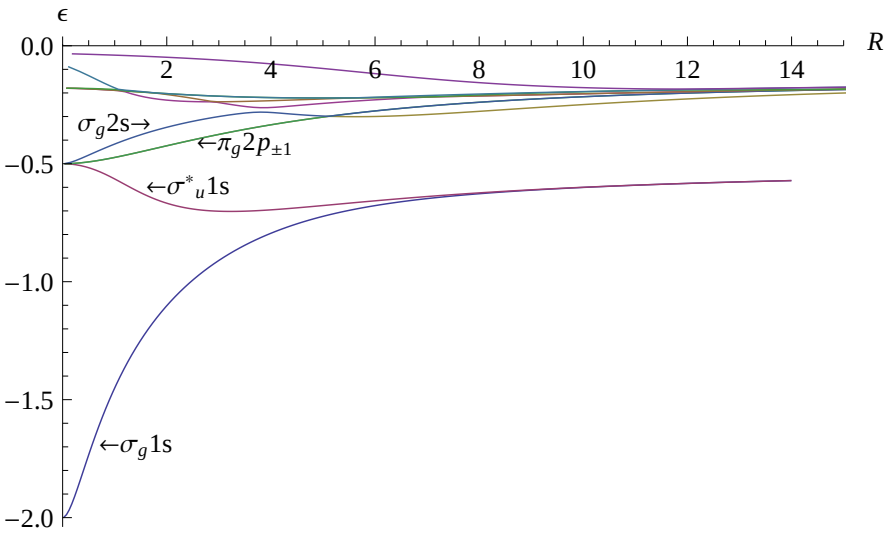


Figure 14.1: Energies of the ground state and excited states of H_2^+, calculated by solving equation (14.1.14) for a basis consisting of 5 Coulomb Sturmians on each nucleus. The energies are expressed in Hartrees and are given as a function of the internuclear separation R, expressed in Bohrs.

If the sum in Equation (14.1.15) is taken over all τ's, the equality holds exactly. When a truncated sum is used, it holds only approximately. With the help of this sum rule, the secular equations (14.1.14) can be rewritten in the form

$$\sum_{\tau}\left[\sum_{\tau''}K_{\tau',\tau''}K_{\tau'',\tau}-kK_{\tau',\tau}\right]C_{\tau,\zeta}=0 \tag{14.1.17}$$

Now suppose that we have solved the secular equation

$$\sum_{\tau}\left[K_{\tau',\tau}-k\delta_{\tau',\tau}\right]C_{\tau,\zeta}=0 \tag{14.1.18}$$

The values of k and $C_{\tau,\zeta}$ thus obtained will also be solutions to (14.1.17). The advantage of using Equation (14.1.18) rather than (14.1.14), i.e. the secular equation involving K rather than $\mathfrak{W}$ and $\mathfrak{S}$, is that for small internuclear separations, the $(\mathfrak{W},\mathfrak{S})$-formulation suffers from problems of over completeness. This is avoided when solving Equation (14.1.18). At large separations, however, the $(\mathfrak{W},\mathfrak{S})$-formulation is accurate, while the K-formulation is less so. We therefore use the K-formulation for small separations, and the $(\mathfrak{W},\mathfrak{S})$-formulation for larger ones, mixing the solutions in a small transition region between the two regimes.

In Figure 14.1, we show the calculated energies of the first few states of the H_2^+ ion, obtained by solving Equation (14.1.14) with a basis consisting of 5 Coulomb Sturmians on each nucleus. The orbitals are labeled using spectroscopic notation for diatomic molecules. For diatomic molecules, where the z-axis is chosen along the internuclear axis, the azimuthal quantum number m is still a good quantum number. That is, only atomic orbitals with the same m enter a particular molecular orbital. In a notation similar to the $s, p, d, f, g, \ldots$ notation for atomic orbitals, one then writes

$\lvert m\rvert$	0	1	2	3	4
Symbol	σ	π	δ	ϕ	γ

See Levine [1991–2009], Chapter 13. For homonuclear diatomic molecules, symmetry or antisymmetry under inversion through the midway point is also a good quantum number. Orbitals that

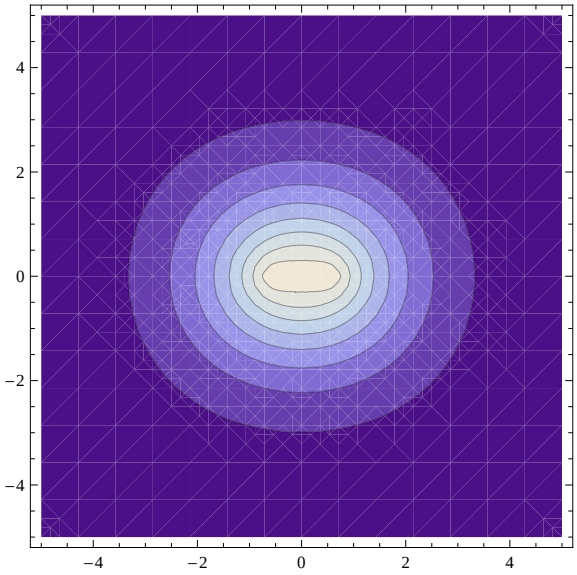


Figure 14.2: The ground state molecular orbital of the H_2^+ ion at nuclear separation $R = 1.21702$ Bohrs ($S = 2, k = 1.64335$). In the united-atom limit, $k = 2$.

are symmetric under inversion are labeled g (gerade), and those that are antisymmetric are labeled u (ungerade). A superscript '$*$' denotes an *anti-bonding* orbital, i.e. one that destabilizes a bond rather than stabilizing it. Finally, the dominant atomic orbital is named in the notation.

Figures 14.2 through 14.5 show the ground state of the H_2^+ ion, i.e. the one-electron, two-center problem with atomic charges $(Z_1, Z_2) = (1, 1)$, for a range of internuclear separations. The figures illustrate the flexibility resulting from the automatic scaling properties intrinsic to the method. The diffuse wave functions correspond to small values of k, while the contracted wave functions correspond to large k. Both situations, as well as the intermediate regions, are handled smoothly using the same small basis set.

In the next chapter, we will show how to construct many-electron configurations from the molecular orbitals discussed above.

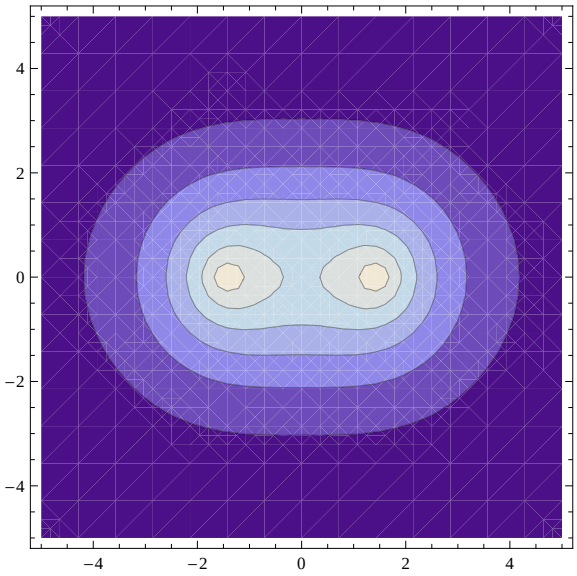


Figure 14.3: The same state at nuclear separation 2.98216 Bohrs $(S = 4, k = 1.34131)$.

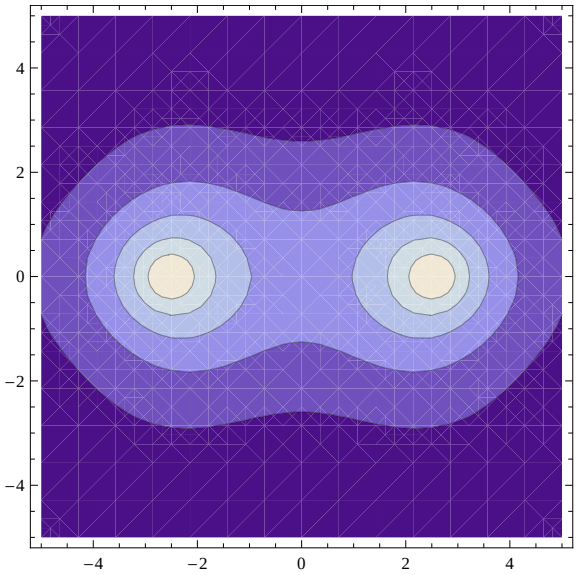


Figure 14.4: Here the internuclear distance has been increased to 5.13325 Bohrs $(S = 6, k = 1.16885)$.

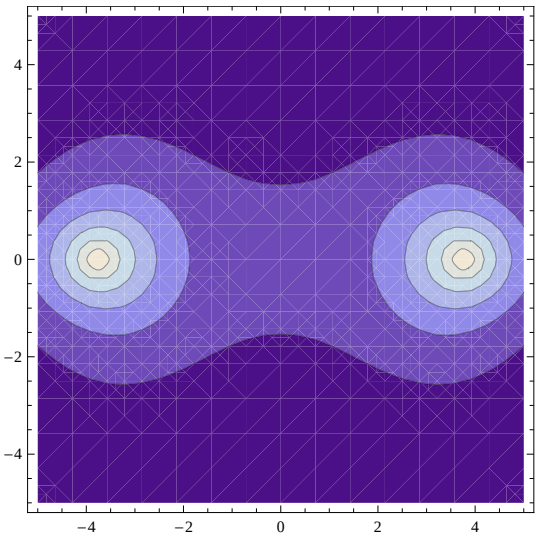


Figure 14.5: The same state with nuclear separation 7.50577 Bohrs $(S = 8, k = 1.06585)$. As the nuclear separation increases, k approaches 1.

CHAPTER 15

N-electron molecular calculations based on Coulomb Sturmians

15.1 Molecular calculations using the isoenergetic configurations

We now introduce the N-electron configurations which are Slater determinants of the form:

$$\Phi_\nu(\mathbf{x}) = |\varphi_{\zeta_1}\varphi_{\zeta_2}...\varphi_{\zeta_N}| \tag{15.1.1}$$

Since the individual molecular orbitals satisfy (14.1.14), the configurations $\Phi_\nu(\mathbf{x})$ are solutions to the separable N-electron equation:

$$\sum_{j=1}^{N}\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(\mathbf{x}_j)\right]\Phi_\nu(\mathbf{x}) = 0 \tag{15.1.2}$$

which can also be written in the form:

$$\left[\sum_{j=1}^{N}\left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right) + \beta_\nu V_0(\mathbf{x})\right]\Phi_\nu(\mathbf{x}) = 0 \tag{15.1.3}$$

where

$$V_0(\mathbf{x}) = \sum_{j=1}^{N} v(\mathbf{x}_j) \quad \text{with} \quad v(\mathbf{x}_j) = -\sum_a \frac{Z_a}{|\mathbf{x}_j - \mathbf{X}_a|} \tag{15.1.4}$$

We would like to use these configurations to build up solutions to the N-electron Schrödinger equation

$$\left[\sum_{j=1}^{N}\left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right) + V(\mathbf{x})\right]\Psi_\kappa(\mathbf{x}) = 0 \tag{15.1.5}$$

with

$$V(\mathbf{x}) = V_0(\mathbf{x}) + V'(\mathbf{x}) = \sum_{j=1}^{N} v(\mathbf{x}_j) + \sum_{i>j}^{N}\sum_{j=1}^{N}\frac{1}{r_{ij}} \tag{15.1.6}$$

and with

$$E_\kappa = -\sum_{j=1}^{N}\frac{k^2}{2} = -\frac{Nk^2}{2} \tag{15.1.7}$$

Thus we write

$$\Psi_\kappa(\mathbf{x}) = \sum_\nu \Phi_\nu(\mathbf{x}) B_{\nu\kappa} \tag{15.1.8}$$

Substituting this into the N-electron Schrödinger equation, and taking the scalar product with a conjugate configuration, we obtain the secular equations:

$$\sum_\nu \int dx\ \Phi^*_{\nu'}(\mathbf{x}) \left[\sum_{j=1}^{N} \left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} \right) + V(\mathbf{x}) \right] \Phi_\nu(\mathbf{x}) B_{\nu\kappa} = 0 \tag{15.1.9}$$

We now introduce a k-independent matrix representing the total potential based on the configurations $\Phi_\nu(\mathbf{x})$:

$$T^{(N)}_{\nu'\nu} \equiv -\frac{1}{k} \int dx\ \Phi^*_{\nu'}(\mathbf{x}) V(\mathbf{x}) \Phi_\nu(\mathbf{x}) \tag{15.1.10}$$

and another k-independent matrix

$$\mathfrak{S}^{(N)}_{\nu'\nu} \equiv \frac{1}{k^2} \int dx\ \Phi^*_{\nu'}(\mathbf{x}) \sum_{j=1}^{N} \left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} \right) \Phi_\nu(\mathbf{x}) \tag{15.1.11}$$

In terms of these matrices, the secular equations become:

$$\sum_\nu \left[T^{(N)}_{\nu'\nu} - k\mathfrak{S}^{(N)}_{\nu'\nu} \right] B_{\nu\kappa} = 0 \tag{15.1.12}$$

Solving the N-electron secular equation (15.1.12), we obtain k for each state κ and thus the energy $E_\kappa = -\frac{Nk^2}{2}$. For a given state κ, the value of k then determines the weighting factors $\beta_{\nu_1}, \beta_{\nu_2}, \ldots$ needed to make each configuration $\Phi_{\nu_1}, \Phi_{\nu_2}, \ldots$ correspond to the same energy E_κ.

In order to build the N-electron matrices $T^{(N)}_{\nu'\nu}$ and $\mathfrak{S}^{(N)}_{\nu'\nu}$ and solve Equation (15.1.12), we must first obtain the coefficients $C_{\tau\zeta}$ by solving (10.5.1) or (10.5.6) from Part II, Section 10.5. In the case of diatomic molecules, we begin by picking a value of the parameter $S = kR$, where R is the interatomic distance and k is the exponent of the Coulomb Sturmian basis set. Neither R nor k is known at this point, but only their product S. As we shall see below, for the diatomic case, all of the integrals involved in Equations (10.5.1) and (10.5.6) are pure functions of S. Having chosen S, we can thus solve the one-electron secular equations and obtain the coefficients $C_{\tau\zeta}$ and the spectrum of ratios k/β_ν. We are then able to solve Equation (15.1.12), which gives us a spectrum of k-values, and thus energies $-Nk^2/2$, and the eigenvectors $B_{\nu\kappa}$. From a k-value, we also get the unscaled distance $R = S/k$. We repeat the procedure for a range of S-values and interpolate to find the solutions as functions of R.

In the case of polyatomic molecules, one can choose a set of angles between the nuclei; these are left fixed under scaling of the coordinate system. The procedure is then similar to that described for the diatomic case.

15.2 Building $T^{(N)}_{\nu'\nu}$ and $\mathfrak{S}^{(N)}_{\nu'\nu}$ from 1-electron components

The matrix $T^{(N)}_{\nu'\nu}$ can be constructed from the 1-electron components. We start with the molecular orbital overlap matrix:

$$\tilde{m}_{\zeta'\zeta} \equiv \int d^3x_j\ \varphi^*_{\zeta'}(\mathbf{x}_j) \varphi_\zeta(\mathbf{x}_j) = \sum_{\tau'} \sum_{\tau} C^*_{\tau'\zeta'} m_{\tau'\tau} C_{\tau\zeta} \tag{15.2.1}$$

where the elements of

$$m_{\tau'\tau} \equiv \int d^3x_j\ \chi^*_{\tau'}(\mathbf{x}_j) \chi_\tau(\mathbf{x}_j) \tag{15.2.2}$$

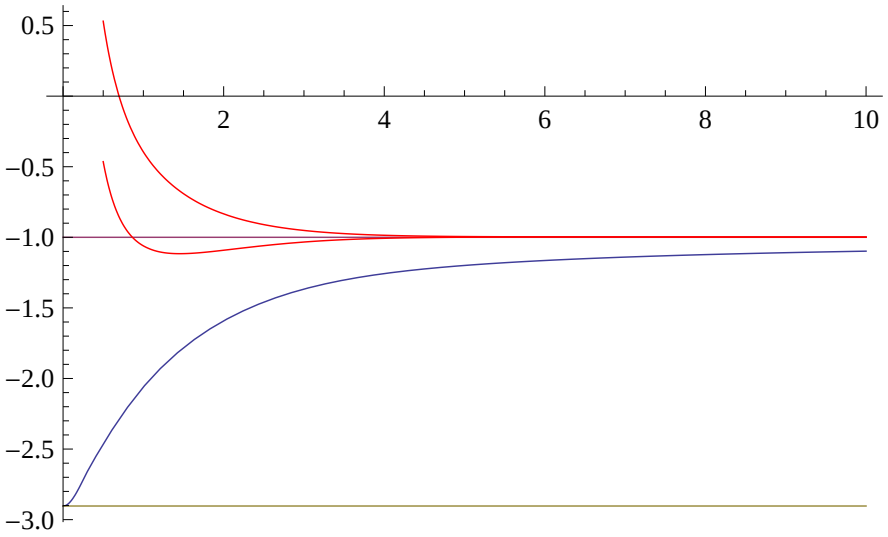


Figure 15.1: This figure shows the results of a preliminary calculation on the dissociation of the hydrogen molecule using a very restricted basis set. Energies are shown in Hartrees as functions of the internuclear separation, measured in Bohrs. The lowest curve shows the ground-state electronic energy by itself, without internuclear repulsion. The two upper curves show the ground state and first excited state electronic energies with nuclear repulsion added, i.e the total energies of the two states. The calculated equilibrium bond length is 1.41 Bohrs, which can be compared with the experimental value, 1.40 Bohrs. It can be seen from the figure that at a separation of 5 Bohrs or more, the molecule is completely dissociated, and in fact the calculated wave function at that internuclear separation corresponds to two neutral hydrogen atoms (Equation (15.2.15)), each with its own electron, while the total energy corresponds to that of two isolated hydrogen atoms.

are displaced Sturmian overlap integrals. The coefficients $C_{\tau\zeta}$ should be normalized in such a way that the diagonal elements of $\tilde{m}_{\zeta'\zeta}$ are all equal to 1. Using the properly normalized coefficients $C_{\tau\zeta}$, we next construct the nuclear attraction matrix based on the molecular orbitals:

$$\tilde{v}_{\zeta'\zeta} \equiv \int d^3x_j \, \varphi^*_{\zeta'}(\mathbf{x}_j) v(\mathbf{x}_j) \varphi_\zeta(\mathbf{x}_j) = -k \sum_{\tau'} \sum_{\tau} C^*_{\tau'\zeta'} \mathfrak{W}_{\tau'\tau} C_{\tau\zeta} \tag{15.2.3}$$

and also

$$\tilde{\mathfrak{S}}_{\zeta'\zeta} \equiv \sum_{\tau'} \sum_{\tau} C^*_{\tau'\zeta'} \mathfrak{S}_{\tau'\tau} C_{\tau\zeta} \tag{15.2.4}$$

The matrices $\tilde{v}_{\zeta'\zeta}$ and $\tilde{\mathfrak{S}}_{\zeta'\zeta}$ are diagonal because the molecular orbitals are solutions to (14.1.14).

Let us now consider a simple illustrative example: the dissociation of the hydrogen molecule, using a very restricted basis set.

Let

$$\begin{aligned}
\varphi_1 &\equiv \varphi_{\sigma_g,1s}(\mathbf{x}) \equiv \frac{1}{\sqrt{2}}\left[\chi_{1s}(\mathbf{x}-\mathbf{X}_1)+\chi_{1s}(\mathbf{x}-\mathbf{X}_2)\right] \equiv \frac{1}{\sqrt{2}}\left[\chi_1+\chi_2\right] \\
\varphi_{\bar{1}} &\equiv \varphi_{\sigma_g,\bar{1}s}(\mathbf{x}) \equiv \frac{1}{\sqrt{2}}\left[\chi_{\bar{1}s}(\mathbf{x}-\mathbf{X}_1)+\chi_{\bar{1}s}(\mathbf{x}-\mathbf{X}_2)\right] \equiv \frac{1}{\sqrt{2}}\left[\chi_{\bar{1}}+\chi_{\bar{2}}\right] \\
\varphi_2 &\equiv \varphi_{\sigma_u^*,1s}(\mathbf{x}) \equiv \frac{1}{\sqrt{2}}\left[\chi_{1s}(\mathbf{x}-\mathbf{X}_1)-\chi_{1s}(\mathbf{x}-\mathbf{X}_2)\right] \equiv \frac{1}{\sqrt{2}}\left[\chi_1-\chi_2\right] \\
\varphi_{\bar{2}} &\equiv \varphi_{\sigma_u^*,\bar{1}s}(\mathbf{x}) \equiv \frac{1}{\sqrt{2}}\left[\chi_{\bar{1}s}(\mathbf{x}-\mathbf{X}_1)-\chi_{\bar{1}s}(\mathbf{x}-\mathbf{X}_2)\right] \equiv \frac{1}{\sqrt{2}}\left[\chi_{\bar{1}}-\chi_{\bar{2}}\right]
\end{aligned} \tag{15.2.5}$$

where $\varphi_{\bar{1}}$ means "$\zeta = 1$, spin down", φ_1 means "$\zeta = 1$, spin up", et cetera. From these, we construct two-electron configurations:

$$\Phi_{1,\bar{1}} \equiv |\varphi_1\varphi_{\bar{1}}| \equiv \frac{1}{\sqrt{2}}\begin{vmatrix} \varphi_1(1) & \varphi_{\bar{1}}(1) \\ \varphi_1(2) & \varphi_{\bar{1}}(2) \end{vmatrix} \tag{15.2.6}$$

and

$$\Phi_{2,\bar{2}} \equiv |\varphi_2\varphi_{\bar{2}}| \equiv \frac{1}{\sqrt{2}}\begin{vmatrix} \varphi_2(1) & \varphi_{\bar{2}}(1) \\ \varphi_2(2) & \varphi_{\bar{2}}(2) \end{vmatrix} \tag{15.2.7}$$

Then

$$\begin{aligned}
\Phi_{1,\bar{1}} &\equiv |\varphi_1\varphi_{\bar{1}}| \\
&= \frac{1}{2}|\left[\chi_1+\chi_2\right]\left[\chi_{\bar{1}}+\chi_{\bar{2}}\right]| \\
&= \frac{1}{2}\left[|\chi_1\chi_{\bar{1}}|+|\chi_1\chi_{\bar{2}}|+|\chi_2\chi_{\bar{1}}|+|\chi_2\chi_{\bar{2}}|\right]
\end{aligned} \tag{15.2.8}$$

and similarly

$$\Phi_{2,\bar{2}} = \frac{1}{2}\left[|\chi_1\chi_{\bar{1}}|-|\chi_1\chi_{\bar{2}}|-|\chi_2\chi_{\bar{1}}|+|\chi_2\chi_{\bar{2}}|\right] \tag{15.2.9}$$

Thus

$$\frac{1}{\sqrt{2}}\left[\Phi_{1,\bar{1}}+\Phi_{2,\bar{2}}\right] = \frac{1}{2\sqrt{2}}\left[|\chi_1\chi_{\bar{1}}|+|\chi_2\chi_{\bar{2}}|\right] \tag{15.2.10}$$

while

$$\frac{1}{\sqrt{2}}\left[\Phi_{1,\bar{1}}-\Phi_{2,\bar{2}}\right] = \frac{1}{2\sqrt{2}}\left[|\chi_1\chi_{\bar{2}}|+|\chi_2\chi_{\bar{1}}|\right] \tag{15.2.11}$$

We can also obtain the relations

$$\begin{aligned}
\Phi_{1,\bar{2}} &\equiv |\varphi_1\varphi_{\bar{2}}| \\
&= \frac{1}{2}|\left[\chi_1+\chi_2\right]\left[\chi_{\bar{1}}-\chi_{\bar{2}}\right]| \\
&= \frac{1}{2}\left[|\chi_1\chi_{\bar{1}}|-|\chi_1\chi_{\bar{2}}|+|\chi_2\chi_{\bar{1}}|-|\chi_2\chi_{\bar{2}}|\right]
\end{aligned} \tag{15.2.12}$$

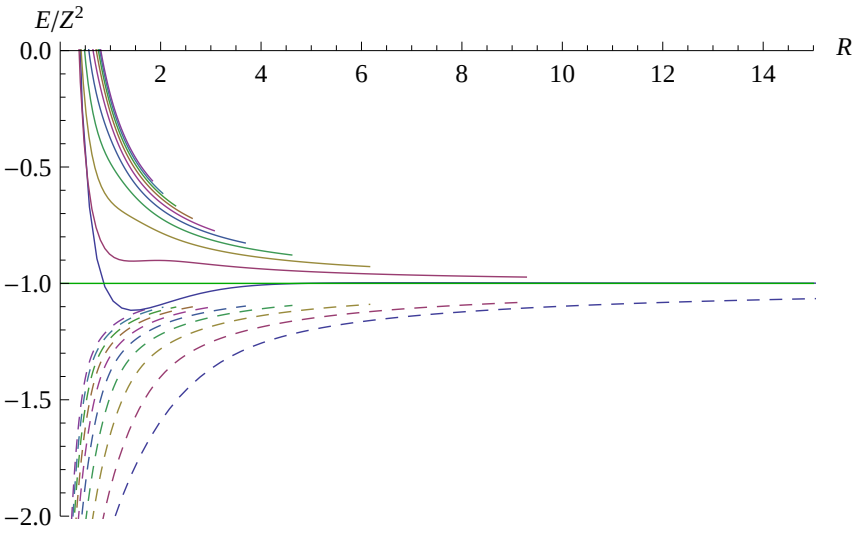


Figure 15.2: This figure shows ground-state energies divided by Z^2 for the 2-electron isoelectronic series for homonuclear diatomic molecules, Z being the nuclear charges. The energies in Hartrees are shown as functions of the interatomic distance R, measured in Bohrs. The dotted curves are electronic energies alone, while the solid curves also include internuclear repulsion. As in Figure 15.1, a very restricted basis set was used for the calculation.

and

$$\begin{aligned}\Phi_{2,\bar{1}} &\equiv |\varphi_2\varphi_{\bar{1}}| \\ &= \frac{1}{2}|\,[\chi_1-\chi_2]\,[\chi_{\bar{1}}+\chi_{\bar{2}}]\,| \\ &= \frac{1}{2}\,[|\chi_1\chi_{\bar{1}}|+|\chi_1\chi_{\bar{2}}|-|\chi_2\chi_{\bar{1}}|-|\chi_2\chi_{\bar{2}}|] \end{aligned} \tag{15.2.13}$$

so that

$$\frac{1}{\sqrt{2}}\left[\Phi_{1,\bar{2}}+\Phi_{2,\bar{1}}\right] = \frac{1}{2\sqrt{2}}\,[|\chi_1\chi_{\bar{1}}|-|\chi_2\chi_{\bar{2}}|] \tag{15.2.14}$$

while

$$\frac{1}{\sqrt{2}}\left[\Phi_{1,\bar{2}}-\Phi_{2,\bar{1}}\right] = \frac{1}{2\sqrt{2}}\,[-|\chi_1\chi_{\bar{2}}|+|\chi_2\chi_{\bar{1}}|] \tag{15.2.15}$$

In the calculation shown in Figure 15.1, the ground state wave function changes character as a function of the internuclear separation, R. As $R \to 0$, the wave function becomes more and more dominated by the configuration $\Phi_{1,\bar{1}}$, which is built from two *gerade* molecular orbitals. But as the molecule dissociates, the wave function becomes the linear combination of configurations shown in equation (15.2.15), which represents two isolated neutral hydrogen atoms, each with its own electron, and the total energy corresponds to that of two isolated hydrogen atoms.

Figure 15.2 shows the energies of the two-electron homonuclear isoelectronic series, i.e. the two-electron diatomics with $Z_1 = Z_2$. The dashed curves show the electronic energies alone, while the solid curves include internuclear repulsion. For both sets of curves, the lower the curve, the smaller the value of Z. It can be seen, that for large values of Z, the curves become closely spaced, and approach a limit as $Z \to \infty$. Our calculation shows not only the well-known bound state for $Z = 1$, but also a meta-stable unbound state for $Z = 2$: While the energy of this state is well above the dissociation limit, there is a shallow local minimum in the total energy curve, that could hold the system together for an extremely short period of time.

15.3 Concluding remarks

Because of the limitations on time, we have as yet only performed very simple pilot calculations. However, the method described in this chapter is general, and potentially highly accurate, when larger basis sets are used. This is because of the automatic scaling properties of molecular Sturmians and because almost all the integrals are evaluated exactly. In particular, behaviour at the cusps of the wave function near nuclei is reproduced accurately, as is long-range behaviour. The automatic scaling of the basis set additionally ensures that the wavefunction is adequately represented, regardless of whether the states are tightly bound or diffuse.

Part IV

Discussion

Discussion

Remarks on the finite element DFT method

It was found that the SIESTA-like DFT method extended with the finite element method gave rapid and accurate results. These results could be tested by comparing them with standard quantum chemical calculations in the case of the isolated molecule, and good agreement was found. The validity of the method for handling effects of an electrostatic environment was investigated by comparing the calculations to a series of approximations, all of which showed that the programs on the whole must be correct.

Comparison with the experimental results in Kubatkin et al. [2003] gave disagreement in many features. In particular, the experimental results showed a spectrum of addition energies in the SET environment that was incompatible with a pure electrostatic effect. Furthermore, the addition energies reported from the experiment showed the molecule in the single electron transistor environment to be extremely volatile, i.e. easily subject to autoionization, with small changes in the environment's work function leading to accepting or donating many electrons. The calculations do not show this feature. It seems likely that in the experiments of Kubatkin et al., the molecule was distorted by its environment, or alternatively, that, due to autoionization, a different part of the spectrum was measured than that of the neutral molecule. Since many tests show the calculational method developed in this thesis to be sound, it would be desirable to compare with other experiments. However, the limitation of time as well as in comparable experiments did not allow this.

Future Work

- The current implementation only uses a no-spin LDA, which is the crudest exchange-correlation functional possible. The method, however, does not depend on this. It should be possible without too much effort to add other XC-functionals to the implementation, using for example `libxc` from the `octopus` project (Castro et al. [2006]). Pure gradient methods like GGA will be especially easy to add, since the piece-wise polynomial representation of finite-element functions make exact differentiation trivial.
- Forces and dynamics have not been implemented, and require some extra work to develop. However, the influence of an electrostatic environment on the molecular geometry may well have a significant effect on the results. Consequently, the use of static geometries that are optimized in vacuum using standard software may constitute a source of error compared to experiments. It is planned to add forces and geometry relaxation to the programs in the near future.
- There is a good deal of room for improvement in the mesh-generating algorithms. Reducing the number of degrees of freedom while retaining accuracy can drastically improve performance. Mesh generation has been restricted to uniform subdivision due to the restrictions posed by the Deal.II-library. However, it is possible to generate mesh structures

directly, allowing us to construct the crude finite element meshes with nodal positions only where it is necessary to represent the atomic orbitals well. The refined mesh will then have significantly fewer points for the same level of accuracy.

- It may be feasible, due to the high level of sparsity arising from the finite element method, to get rid completely of the atomic orbital basis functions and simply use the finite element basis all the way through calculations. This will yield results that are not biased due to the atomic orbital shapes. It also gets rid of the need to build real-space representations of the basis functions, a step that requires either $\mathcal{O}(N_{\text{orbs}}N_q)$ storage, or a heavy $\mathcal{O}(N_{\text{orbs}}N_q)$ computation in each SCF step.

- A posteriori error estimates should be developed and implemented. Besides improving the adaptive mesh refinement, this will allow the program to certify that a solution is within a given error bound.

- Besides electronic nano-devices, the method might be applied to systems in biology, where part of the system is quantum-active but where this part is embedded in a larger environment that modifies its behaviour. It has been recommended to investigate the redox reactions that take place in proteins like hemoglobin. The quantum-active part of hemoglobin is the iron-porphyrin group, but the behaviour of this group is modified by the amino-acid chain that surrounds it. It may well be possible to use the finite-element DFT-method developed in Part I to give a good yet efficient description of the these biological redox reactions, which are closely analogous to the redox reactions that take place in molecular single-electron transistors. Other biological applications might include calculations on the cytochrome-C molecule and other components of the electron transport chain.

Methods for evaluating interelectron repulsion integrals for ETOs

This is a difficult mathematical problem on which a number of gifted authors have worked during the last 60 years. Some progress has been made in this field, but currently available computer programs remain very slow. In this thesis, a new method is developed specifically for molecular orbitals based on Coulomb Sturmians. This method uses the Fourier transform form of the interelectron repulsion and also makes use of the special properties of Coulomb Sturmians. In particular, it makes use of the Fock projection, which relates the Fourier transform of Coulomb Sturmian basis functions to hyperspherical harmonics.

A chapter is devoted to the mathematical methods employed, since these are novel and central to the evaluation. In particular, the properties of harmonic polynomials and hyperspherical harmonics are discussed. These properties lead to powerful theorems for the evaluation of angular and hyperangular integrals. Much of the speed and convenience of the method discussed in this part of the thesis is due to the ease with which angular and hyperangular integrals can be evaluated.

A second mathematical section is devoted to the Legendre polynomial expansion of displaced functions. Even the difficult three- and four-center interelectron repulsion integrals for ETOs can evaluated approximately by this method, but the method involves a series that must be truncated at some point. Because of the limitations of time, we were not able to perform sufficiently accurate calculations. Further work on this method is needed to determine whether the method can achieve satisfactory accuracy while maintaining efficiency.

An alternative to the Legendre polynomial expansion is also discussed. This is a special version of the Gaussian expansion method especially adapted to Coulomb Sturmians. The method has the special feature that scaling is automatic, and hence universal expansion coefficients can be used for both contracted and diffuse orbitals. This allows us to perform the integrals once and

for all, and to store exact functions expressing them. The heavy work can be done off line, and the results used in all situations. The method gives a very satisfactory accuracy.

The most important ETO interelectron repulsion integrals are those involving only two centers, with each charge density localized on one center. These integrals have the greatest magnitude and extend to large internuclear separations. They can be evaluated rapidly and exactly using the properties of hyperspherical harmonics.

Mixed density interelectron repulsion integrals are smaller in magnitude and fall off rapidly after a moderate internuclear separation. As the overlap integrals between orbitals vanish, so do the mixed-density integrals. Because of this, the error due to the asymptotic behaviour of Gaussian orbital expansions becomes unimportant.

In conclusion the hyperspherical method, applied to interelectron repulsion integrals involving molecular Sturmians, offers a rapid and convenient method for their evaluation. The most important integrals are evaluated exactly, and specially adapted Gaussian expansions are applied only to the less important integrals involving mixed densities. The special Gaussian expansions of Coulomb Sturmians make use of universal expansion coefficients and automatic scaling, so that the heavy work can be done off line. The results can be used in all situations, and thus the overall hyperspherical method has wide applicability. The method has been completely implemented in the case of diatomic molecules, and partially implemented in the general case.

Future Work

- The `C++`-library and the automatic `C`-code generating facilities are a work in progress and need further development to reach a mature enough state for public consumption. Preliminary results, however, look extremely promising, and I will be dedicating a great deal of work on this over the course of the coming year.

- The methods described in Chapter 12 can be used to synthesize other types of atomic orbitals than Coulomb Sturmians. The crucial point is that we can precalculate the expansion coefficients for the radial functions, which in turn lets us carry out all of the heavy calculations once and for all off-line. In particular, interelectron repulsion integrals for the numerical atomic orbitals used in the DFT method in Part I could be treated in this fashion. Similar to the case of Coulomb Sturmians, we can compute interelectron repulsion integrals directly by way of the spherical Bessel Transform in the pure-density case. Efficient and accurate methods for computing this transform for numerical functions with finite support can be found for example in Talman [1993], as well as in later works building on Talman's paper. While the numerical orbitals are not automatically scaled, they are precomputed off-line from atomic pseudopotentials. Once computed, they do not change, which allows us to precompute the expansion coefficients for a highly detailed Gaussian expansion to use for the integrals involving mixed-center densities. This in turn lets us carry through the bulk of calculations off-line, as was done in Chapter 12. Thus, as was the case for the Coulomb Sturmians, we should be able to obtain accurate results, while offloading all the heavy work to precalculation. Fast interelectron repulsion integrals for the numerical atomic orbitals will allow us to add a number of hybrid exchange-correlation functionals that involve exact exchange, for example B^3LYP.

The generalized Sturmian method applied to molecules

The generalized Sturmian method has been applied with great success to calculation of atomic spectra (Avery [2008a], Avery and Avery [2005, 2006, 2008], et cetera). In this thesis, a method is developed for applying a closely analogous generalized Sturmian method to molecular calcu-

lations. Like the atomic generalized Sturmian method, the molecular method uses isoenergetic configurations based on solutions to an approximate many-electron Schrödinger equation with a scaled zero-order potential. The scaling factor is chosen in such a way as to make all of the configurations in a basis set correspond to the energy of the state being described. As in the atomic case, this leads to automatic scaling, so that the tightly bound ground state and the progressively diffuse excited states are equally well described by the same basis. The methods for evaluating interelectron repulsion integrals based on molecular Sturmians have the same automatic scaling feature, and thus form a natural part of this scheme. We obtain good solutions to many states at once, using only a very small basis and without any need for SCF calculations. The wave function is in a form that is easy to interpret by inspection or to analyze automatically by computer (see, for example, Tutorial 2 at Avery [2006] for automatic classification of wave functions).

Because of the limitations of time, only a few very simple examples have been treated, but the outline of the method for solving the general cases is complete, and most of it has been implemented. The results of the pilot calculations performed in this thesis are very promising, and it seems likely that the generalized Sturmian method applied to molecules will lead to calculations much more accurate than those based on Gaussians, the basis sets being far smaller but far more flexible and appropriate for synthesis of the actual wave functions.

Future Work

Once we have found a generalized Sturmian basis that converges well, most of the standard techniques in quantum chemistry can be employed in the same way that they are currently used with bases obtained from initial Hartree-Fock calculations. Two obvious steps are to use the frozen core approximation to factor out correlation of core electrons, and use standard perturbation theory based techniques to reduce the computational efforts necessary for configuration interaction. Using the generalized Sturmian method with for example Coupled Cluster methods requires some work, but may be well worth the effort due to improved convergence properties compared to using Hartree-Fock based configurations. In Avery and Avery [2003], we began work on a second-quantized formulation of atomic generalized Sturmians of the Goscinskian type, and this can be extended to include the molecular Sturmians developed here. This is the first step towards developing a Sturmian Coupled Cluster method.

Both generalized Sturmians of the Goscinskian type (See e.g. Avery and Avery [2006]), and isoenergetic configurations of Coulomb Sturmians, work well for accurately and rapidly calculating many states of few-electron systems, using only very few basis functions. However, as the number of electrons grows, $V_0(\mathbf{x}) = \sum_I \sum_{j=1}^{N} |\mathbf{x} - \mathbf{X}_I|^{-1}$ less and less resembles the full potential V, and so the derived one-electron molecular orbitals less efficiently represent the true wave functions. By incorporating correlation effects into V_0, we can build basis sets that are better suited to deliver compact representations of many-electron states. Some interesting variations are under way. For example, Professor Gustavo Gasaneo and his students at Universidad Nacional del Sur in Argentina have worked on numerical Sturmians that incorporate interelectron repulsion effects.

In addition, we have not yet investigated the generalized Sturmian method for systems that look very different from atoms and molecules, although there is nothing inherent in the general method that restricts us to these systems, nor even to electronic structure calculations.

Aside from these more structural and long term improvements, there are some simpler directions of further work that can be pursued immediately:

- **Separating out the core**: We can take advantage of the fact that the valence electrons can, to some approximation, be treated separately from the closed shells: An electron in a closed

shell interacts only slightly with those in other shells. One approximation that takes advantage of this is the *frozen core approximation*.

Treating only valence electrons as being free greatly curbs the steep growth in the degrees of freedom when the number of electrons is increased, and is likely to be absolutely necessary to implement if systems with large numbers of electrons are to be treated with reasonable accuracy. Standard methods exist that possibly can be applied directly to the generalized Sturmian method.

- **Incremental computation**: Using creation and annihilation operators, it is possible to express the $N+1$-electron integrals in terms of N-electron integrals. On this basis, it is possible to to create an incremental algorithm that uses smaller systems to solve larger ones.

- **Automatic generation of symmetry adapted basis sets**: When calculating states of a particular symmetry, only a select set of basis functions contribute. Version 1.0 of the Generalized Sturmian Library (Avery [2008b]) includes three prototypes for automatically and systematically generating basis sets that contain only basis vectors that contribute to a particular symmetry. It is planned to extend this work in the near future.

Index

Bibliography

J. Avery and P. J. Ørmen. Transferable integrals in a deformation density approach to molecular orbital calculations; iv. evaluation of angular integrals by a vector-pairing method. *Int. J. Quantum Chem.*, 18:953–971, 1980. 79, 86, 87, 98

J. E. Avery. libsturm: The generalized sturmian library. `http://sturmian.kvante.org`, 2006. 150

J. E. Avery. The generalized Sturmian method: Development, implementation, and applications in atomic physics. Master's thesis, Department of Computer Science, University of Copenhagen (DIKU), October 2008a. `http://sturmian.kvante.org/thesis/speciale.pdf` and `http://sturmian.kvante.org/thesis/sturmian-1.0.pdf`. 4, 149

J. E. Avery. *The Generalized Sturmian Library, version 1.0*. Dep. of Computer Science, University of Copenhagen (DIKU), `http://sturmian.kvante.org`, September 2008b. (Second part of the master's thesis). 151

J. E. Avery and J. S. Avery. Autoionizing states of atoms calculated using generalized sturmians. *Advances in Quantum Chemisty*, 49:103–118, 2005. 149

J. E. Avery and J. S. Avery. *Generalized Sturmians and Atomic Spectra*. World Scientific, December 2006. ISBN 981-256-806-9. ISBN: 981-256-806-9. iii, 4, 90, 149, 150

J. E. Avery and J. S. Avery. Atomic core-ionization energies; approximately piecewise-linear and linear relationships. *Journal of Mathematical Chemistry*, 2008. 149, 215

J. E. Avery, L. Kristiansen, and J.-Y. Moyen. Static complexity analysis of higher order programs. In *FOPARA'09 Proceedings of the First international conference on Foundational and practical aspects of resource analysis*. Springer Verlag, 2010. 235

J. S. Avery. *Hyperspherical Harmonics*. Kluwer Academic Publishers, 1989. 78

J. S. Avery and J. E. Avery. Kramers pairs in configuration interaction. *Advances in Quantum Chemistry*, 43:185–206, 2003. 150

J. S. Avery and J. E. Avery. Can Coulomb Sturmians be used as a basis for n-electron molecular calculations? *J. Phys. Chem. A*, 113(52), 2009. 206

W. Bangerth, R. Hartmann, and G. Kanschat. deal.II — a general-purpose object-oriented finite element library. *ACM Trans. Math. Softw.*, 33(4). `http://www.dealii.org`. 14, 15, 48

S. Brenner and L. Scott. *The mathematical theory of finite element methods*. Texts in applied mathematics. Springer-Verlag, third edition, 2008. 9, 13, 20

S. C. Brenner and C. Carstensen. *Encyclopedia of Computational Mechanics*, volume 1, chapter Chapter 4, Finite Element Methods. John Wiley & Sons, 2004. 13, 20

E. J. Bylaska, M. Holst, and J. H. Weare. Adaptive finite-element method for solving the exact kohn-sham equation of density functional theory. *J. of Chemical Theory and Computation*, 2009. 8

A. Castro, H. Appel, M. Oliveira, C. Rozzi, X. Andrade, F. Lorenzen, M. Marques, E. Gross, and A. Rubio. octopus: a tool for the application of time-dependent density functional theory. *Phys. Stat. Sol. B*, 243:2465–2488, 2006. 147

P. A. Dirac. Quantum mechanics of many electron systems. *Proceedings of the Royal Society of London. Series A*, pages 714–733, April 1929. 1

V. Fock. 1935. *Z. Phys.*, 98:145, 1935. 76

M. J. Frisch, G. W. Trucks, H. B. Schlegel, and G. E. S. et al. Gaussian 03, Revision B.05, 2003. Gaussian, Inc., Pittsburg, PA, 2003. 54

M. Galassi. *GNU Scientific Library Reference Manual*, 3rd edition. ISBN 0954612078. 126

J. Heinze, J. Mortensen, K. Müllen, and R. Schenk. The charge storage mechanism of conducting polymers: A voltammetric study on defined soluble oligomers of the penylene-vinylene type. *J. Chem. Soc. Chem. Commun.*, pages 701–703, 1987. 64

P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, Nov 1964. doi: 10.1103/PhysRev.136.B864. 25, 26

J. F. Janak. Proof that $\partial e/\partial n_i = \epsilon_i$ in density-functional theory. *Physical Review B*, 18(12):7166–7168, December 1978. 27, 59

B. Jenkins. Minimal perfect hashing. http://burtleburtle.net/bob/hash/perfect.html, 1996–2006. 128

C. Johnson. *Numerical Solutions of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, 1987. 9

K. Kaasbjerg and K. Flensberg. Strong polarization-induced reduction of addition energies in single-molecule nanojunctions. *Nano Letters*, 8(11):3809–3814, October 2008. DOI: 10.1021/nl8021708. 3, 8

B. S. Kirk, J. W. Peterson, R. H. Stogner, and G. F. Carey. `libMesh`: A C++ Library for Parallel Adaptive Mesh Refinement/Coarsening Simulations. *Engineering with Computers*, 22(3–4):237–254, 2006. `http://libmesh.sourceforge.net`. 48, 49

L. Kleinman and D. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48: 1425–1428, 1982. 30

T. Koga and T. Matsuhashi. One-electron diatomics in momentum space. v. nonvariational lcao approach. *J. Chem. Phys.*, 89:983, 1988. 134

W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, Nov 1965. doi: 10.1103/PhysRev.140.A1133. 26

T. Koopmans. über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelend Elektronen eines Atoms. *Physica*, 1(1-6):104–113, 1934. 58

S. Kubatkin, A. Danilov, M. Hjort, J. Cornil, J.-L. Bredas, N. Stuhr-Hansen, P. Hedegard, and T. Bjornholm. Single-electron transistor of a single organic molecule with access to several redox states. *Nature*, 425(6959):698–701, 2003. doi: 10.1038/nature02010. URL `http://dx.doi.org/10.1038/nature02010`. 63, 64, 65, 68, 147

I. N. Levine. *Quantum Chemistry*. Pearson Education, Inc., 6 edition, 1991–2009. 135

A. Logg and G. Wells. Dolfin: Automated finite element computing. *ACM Transactions on Mathematical Software*, 37(2), 2010. `http://www.fenicsproject.org`. 49

J. Luo, Z. Q. Xue, W. M. Lui, J. L. Wu, and Z. Q. Yang. Koopmans' theorem for large molecular systems within density functional theory. *Journal of Physical Chemistry A*, 2006(110):12005–12009, 2006. 27, 59

M. Michels. Evaluation of angular integrals in rotational invariants. *Int. J. Quantum Chem.*, 20: 951–952, 1981. 79, 86, 87, 98

NIST CCCBDB. Nist computational chemistry comparison and benchmark database: Nist standard reference database number 101. http://cccbdb.nist.gov/, April 2010. Release 15a. 54

V. Papaefthimiou, S. Siokou, and S. Kennou. Growth and interfacial studies of conjugated oligomer films on si and sio_2 substrates. *Journal of Applied Physics*, 91(7):4213–4219, April 2002. 54

J. Pask and P. Sterne. Finite element methods in ab initio electronic structure calculations. *Modelling and Simulation in Materials Science and Engineering*, 13:R71–R96, 2005. 8

J. P. Perdew and M. Levy. Comment on "significance of the highest occupied kohn-sham eigenvalue". *Physical Review B*, 56(24):16021–16028, December 1997. 59

P. Politzer and F. Abu-Awwad. A comparative analysis of hartree-fock and kohn-sham orbital energies. *Theoretical Chemistry Accounts*, 99:83–87, 1998. 27, 59

P. Popelier, editor. *Solving the Schrödinger Equation: Has Everything been Tried*. Imperial College Press, 2011. 171

J. C. Rivière. The work function of gold. *Applied Physics Letters*, 8(7):172, April 1966. 56, 66

T. Shibuya and C. Wulfman. Molecular orbitals in momentum space. *Proc. Roy. Soc. A*, 286:376, 1965. 89

J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal. The siesta method for ab initio order- n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002. URL `http://stacks.iop.org/0953-8984/14/i=11/a=302`. 3, 7, 29, 30, 32, 33

P. Sommer Larsen. *Ph.D. Thesis*. PhD thesis, Chemistry Department, University of Copenhagen, Ph.D. Thesis. 94

J. D. Talman. Molecular scf calculations using a basis of numerical orbitals. *International Journal of Quantum Chemistry*, 27:321–330, 1993. 149

N. Toda, H. Kimizuka, and S. Ogata. DFT-based FEM analysis of nonlinear effects on indentation process in diamond crystal. *International Journal of Mechanical Sciences*, 52:303–308, 2009. 8

R. van Leeuwen. *Kohn-Sham Potentials in Density Functional Theory*. PhD thesis, Vrije Universiteit, Amsterdam, 1994. 25

Publications on the new FEM+DFT-method

[DRAFT]: Finite-element DFT Calculations for a Single Molecule, Single Electron Transistor

Finite-element DFT Calculations for a Single Molecule, Single Electron Transistor (Unfinished Draft)

James Avery*
Dep. of Computer Science, University of Copenhagen (DIKU), Denmark

Stig Skelboe†
Dep. of Computer Science, University of Copenhagen (DIKU), Denmark

Kristen Kaasbjerg‡
Dep. of Physics, CAMD, Technical University of Denmark (DTU), Denmark

Mattias Lein§
Centre for Theoretical Chemistry and Physics, New Zealand Institute for Advanced Study, Massey Universy, Auckland, New Zealand

Kurt Stokbro¶
Quantumwise A/S, Lersø Parkallé 107, DK-2100 Copenhagen Ø, Denmark

We present ab initio calculations for a single-electron transistor (SET) that uses a single OPV5-tBu molecule for capacitor. The calculations were done using a newly developed finite-element DFT method that allows efficient treatment of complex molecules interacting with arbitrarily large electrostatic environments. The calculated properties are compared to experiments performed by Kubatkin et al., who have constructed an SET similar to the one used in our calculations. Quite different properties were found by Kubatkin et al.; we point to a number of possible reasons for the discrepancy. Further study is necessary to determine which of these apply.

I. INTRODUCTION

Over the past decade the field of single-molecule electronics[1] has undergone an enormous evolution. Experimentally, three-terminal[2–6] nano-junctions containing a single or few molecules can provide detailed information about the molecular excitations by monitoring the current as function of bias and gate voltage. Typically, such studies are summarized in so-called charge stability diagrams which plot the differential conductance dI/dV as a function of gate and source-drain voltage. Here, molecular excitations show up as peaks in dI/dV at bias values corresponding to the energy spacing between the molecular states. Such devices therefore function as a nice tool for doing bias spectroscopy on single molecules.

Apart from interacting with the internal degrees of freedom of the molecule, the current carrying electrons also interacts with the polarizable environment of the nano-junction via the long-ranged Coulomb interaction. Electron tunneling to and from the molecule therefore results in the formation of a polarization charge in the nearby metallic electrodes and gate oxide. For simple geometries, such as an infinite metallic surface, the induced potential from the polarization charge can be described in terms of image charges inside the surface. Both experiments[3,6] and theoretical descriptions[7] has demonstrated that the interaction with the image charges results in drastically reduced charging energies of the molecule. For isolated molecules typical charging energies, corresponding to the distance in energy between the ionization energy and electron affinity, is on the order of several electron-volts. As the gate-coupling in single-molecule junction is often very low (~ 0.1)[7,8], the image charge effect helps to make more charge states of the molecule experimentally accessible which results in the appearance of full diamonds in the stability diagram[3,5]. When this is the case, a strongly varying size of the diamonds is typically a fingerprint of the different molecular charge states.

In this work, we have developed a numerical method that can calculate charging energies for molecules placed in nano-junctions. It is based on a finite element implementation of DFT that allows us to solve Poisson's equation for the potential in arbitrary junction geometries and with a spatially varying dielectric constant. Such calculations naturally involves an atomic description of the molecule as provided by DFT together with a realistic three-dimensional model of the nano-junction in which Poisson's equation must be solved. The implementation can be used as a tool to analyze experimental results. For example, the size of the Coulomb diamonds inferred from experiments can be calculated. This can help to confirm that the Coulomb diamonds indeed originate from transport through the molecule. Furthermore, the orientation of the molecule relative to the electrodes and gate oxide may be revealed by studying the energetic position of the charge states as a function of applied voltages. Calculations on an OPV5 based single-molecule junction indeed show that the relative sizes of the Coulomb diamonds are in good agreement with experiment[3]. However, the absolute values are much smaller in the experimental stability diagram. Reasons for this discrepancy are discussed in

Section VII.

II. METHOD

Conventional DFT-methods generally scale as $\mathcal{O}\left(N^3\right)$ in the size of the basis set. Although DFT methods are among the fastest ab initio quantum chemical approaches, the growth in resource consumption prohibits practical calculations of medium to large nanosystems. However, there has been a great deal of progress in the last decade in the development linearly scaling DFT-methods. One such order-N method is the SIESTA method (*Spanish Initiative for Electronic Simulations with Thousands of Atoms*), described by Soler and coworkers in[9].

The SIESTA method is based on linear combinations of atomic orbitals (LCAO). The orbitals are localized numerical basis functions optimized for (norm conserving) Troullier-Martins pseudopotentials that yield solutions non-zero only in a small region of space. A compensation charge at each atomic site screens electrostatic interactions. This results in very sparse equations that can be solved with $\mathcal{O}\left(N\right)$ time and space for many classes of systems. The method as presented by Soler et al. relies on representing potentials and the electron density on a uniform real space grid. Because memory requirements for a uniform grid scales as $\mathcal{O}\left(L^3\right)$ in the grid's diameter, this scheme is efficient only when the region of space represented by the grid is small compared to the number of electrons, or rather, when the electron density is spread out more or less evenly over the represented region. This is generally the case when solving for molecules or crystals in vacuum.

However, we wish to be able to treat nano-devices for which a molecule, or a similar nano-structure that must be treated with high accuracy, resides in a macroscopic device that can be treated at low detail. We have developed an extension to the SIESTA method that lifts the real space representation from a regular grid to finite element spaces. This allows multiple levels of detail in treating the system as a whole: The regions where little interesting quantum chemistry goes on can be treated with classical electrostatics using crude representations, while at the same time the quantum chemical parts are treated with high accuracy.

The interaction of the quantum system with the electrostatic environment is based on a model developed by Kristen Kaasbjerg[10], developed for the semi-empirical Hückel method. In this model, the total energy function is extended to include interactions with a number of dielectric and metallic regions surrounding the molecular system.

In this paper, we will not treat the developed method in detail. Rather, a complete description of the finite element DFT method will be published in a separate paper.[11]

III. THE SINGLE-ELECTRON TRANSISTOR ENVIRONMENT

We model the SET environment in the following manner: The molecule rests on top of a dielectric of thickness H and dieclectric constant ϵ_r. The lowest lying nucleus is at distance d_y from the dielectric. On the left and right side of the molecule, at distance d_x from the left-most and rightmost nuclei, are source and drain electrodes, modeled as metallic regions with potential $V_s = V_{sd}/2$ and $V_d = -V_{sd}/2$. The lower surface of the dielectric is enforced a potential V_g.

Except in Section V B 1, the calculations in this paper use $d_x = d_y = 1$Å, $H = 50$Å, and a dielecric constant $\epsilon_r = 10$, corresponding to Al_2O_3. Figure 1 shows the setup.

We use an OPV5 moleculule[12] with tert-butyl groups at the terminals to insulate from strong interaction with the source-drain electrodes. Treating the core electrons with Troullier-Martins pseudo-potentials, we correllate 242 of the 354 electrons. The molecular structure was optimized using[13] with [?] and the PW91 exchange-correlation functional. The structure for the neutral molecule in vacuum was used for all charge states. Further development of our software is needed in order to optimize the geometry in the SET environment.

IV. VALIDITY OF OUR METHOD; ENERGIES IN VACUUM

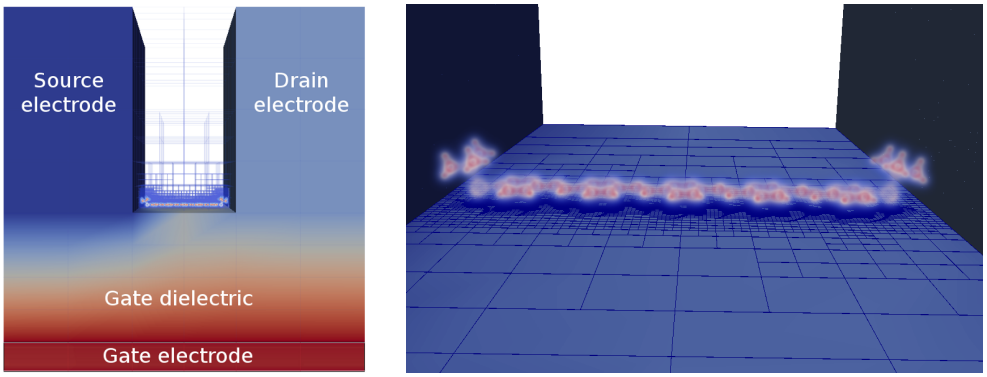
While we are not aware of other software that can treat molecules the size of OPV5-tBu in a large electrostatic environment, we can obtain results for the *isolated* molecule using standard quantum chemical software. This allows us to assess the quality of our calculations: both as a check for errors, and to determine whether LDA is accurate enough to treat this molecule. Reference calculations were made using `Gaussian '03`[13], using a polarized double-zeta basis set, `cc-pVDZ`, and the B^3LYP exchange-correlation functional.

Figure 2 and Table I shows qualitative agreement between the `Gaussian` B^3LYP results and our FEM+LDA calculations. To the extent that the OPV5-tBu molecule can be properly described by B^3LYP, our finite element LDA achieves this as well.

V. RESULTS

A. Total energies, charging energies, and addition energies

Table II shows total energies relative to the neutral molecule as well as the charging energies $\Delta E^q = E^{q+1} - E^q$, i.e. the vertical ionization energies and electron affinities. Values are computed for the isolated molecule



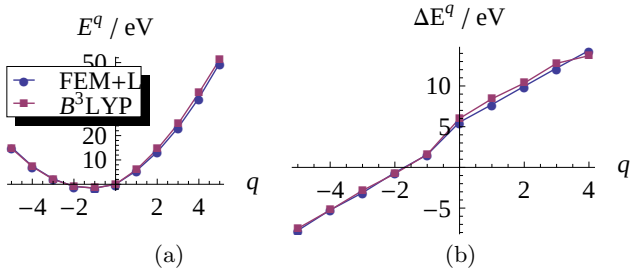
(a) (b)

FIG. 1. (a) OPV5-tBu molecule in a single electron transistor environment, coloured by effective potential $V^{eff}(x)$. Here shown with $V_g = 3V$ and $V_{sd} = 1V$. (b) Close-up view of the molecule, showing the electron density $\rho(x)$. Note the extremely high level of detail in the region where $\rho(x)$ is large.

	B^{3}LYP	FEM+LDA[a]
ΔE^{-5} (5^{th} EA)	−7.50	−7.75
ΔE^{-4} (4^{th} EA)	−5.24	−5.24
ΔE^{-3} (3^{rd} EA)	−2.83	−3.12
ΔE^{-2} (2^{nd} EA)	−0.78	−0.644
ΔE^{-1} (1^{st} EA)	1.60	1.56
ΔE^{0} (1^{st} IP)	6.01	5.55
ΔE^{+1} (2^{nd} IP)	8.43	7.71
ΔE^{+2} (3^{rd} IP)	10.3	9.96
ΔE^{+3} (4^{th} IP)	12.8	12.2
ΔE^{+4} (5^{th} IP)	13.7	14.3

[a] Isolated molecule

TABLE I. Charging energies in eV for the OPV5-tBu molecule, isolated and in the SET-environment. B^{3}LYP-values were computed using `Gaussian '03`, as is discussed in the text.



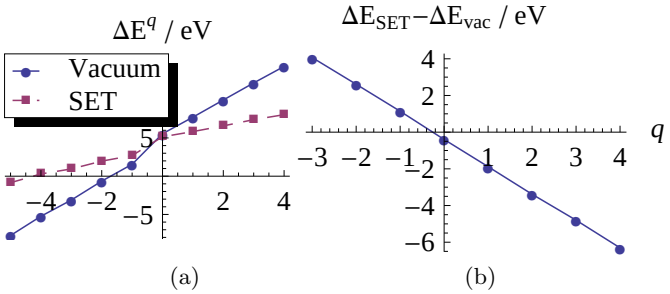
(a) (b)

FIG. 2. Comparison to `Gaussian '03` calculations of OPV5-tBu using B^{3}LYP. (a) Total energy difference from the neutral molecule as a function of the added charge q. (b) Vertical charging energies $\Delta E^q = E^{q+1} - E^q$.

in vacuum and for the molecule placed in the SET environment as described in the previous section. An experimental study by Papaefthimiou et al.[14] found the ionization energy of Ooct-OPV5 (which is expected to have similar electrical properties to OPV5-tBu) to be 5.45 ± 0.1eV, in approximate agreement with our results for OPV5-tBu. The 2% difference may be caused by the somewhat larger size of the Ooct-OPV5 molecule resulting in slightly smaller charging energies; however part of the difference may be due to calculational inaccuracy. A test calculation using our method for the isolated benzene molecule yielded the ionization energy to 1% of the experimental value[15] (9.15 eV versus 9.24 eV). In sum-

	In vacuum	In SET
ΔE^{-5} (5^{th} EA)	−7.75	−0.849
ΔE^{-4} (4^{th} EA)	−5.24	0.293
ΔE^{-3} (3^{rd} EA)	−3.12	0.951
ΔE^{-2} (2^{nd} EA)	−0.644	1.97
ΔE^{-1} (1^{st} EA)	1.56	2.70
ΔE^{0} (1^{st} IP)	5.55	5.15
ΔE^{+1} (2^{nd} IP)	7.71	5.83
ΔE^{+2} (3^{rd} IP)	9.96	6.57
ΔE^{+3} (4^{th} IP)	12.2	7.39
ΔE^{+4} (5^{th} IP)	14.3	8.04

TABLE II. Calculated charging energies (eV) in for the OPV5-tBu molecule in the SET environment compared to the isolated molecule in vacuum.



(a) (b)

FIG. 3. (a) Calculated charging energies for the OPV5-tBu molecule in the SET environment compared to the isolated molecule. (b) Reduction in charging energies from vacuum to the SET environment.

q	−5	−4	−3	−2	−1	0	+1	+2	+3
$\Delta^2 E^q_{\text{Vac}}$	2.52	2.12	2.47	2.2	3.99	2.16	2.25	2.23	2.16
$\Delta^2 E^q_{\text{SET}}$	1.14	0.66	1.02	0.73	2.45	0.68	0.742	0.816	0.65

TABLE III. Calculated charging energies (eV) in for the OPV5-tBu molecule in the SET environment compared to the isolated molecule in vacuum.

mary, we expect the calculated charging energies to be accurate to within a couple of percent.

We note from Table II that the effect of placing the OPV5-tBu molecule in the SET environment is a reduction in the *differences* between charging energies. In fact, as we can see in Figure 3, this reduction is linear in the charge q with $\Delta E^q_{\text{SET}} \approx \Delta E^q_{\text{Vac}} - (1.48q + 0.39)$eV ($\pm$0.1eV). This linear reduction of the charging energies corresponds to an approximately constant reduction of the molecular *addition energies* $\Delta^2 E^q$ of roughly 1.46eV. Table III and Figure 4 shows the addition energies of the molecule in the SET compared to the isolated molecule. In Section VI 3, we explain the approximately linear reduction of the charging energies, and the corresponding roughly constant reduction of $\Delta^2 E^q$.

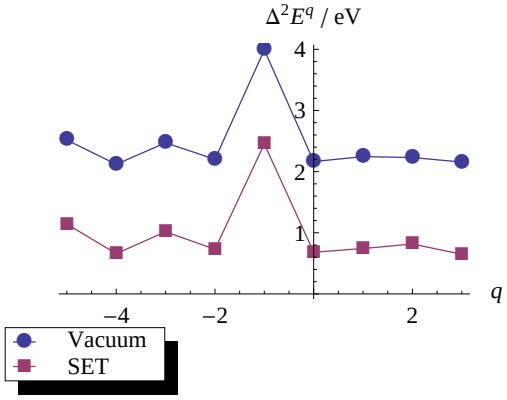


FIG. 4. Calculated addition energies for the isolated OPV5-tBu molecule in vacuum, compared to the molecule in the SET environment.

B. The effect of applying a gate potential

Figure 5 shows total energy curves of the charge states at zero source-drain voltage as a function of the ground voltage V_g. The energies are corrected for the work function for gold, chosen to be $\mathsf{W}_{\text{Au}} = 5.28$eV[16]. The lowest line for a given V_g is thus the most stable charge state for that voltage when the molecule in sin contact with a gold electron reservoir. The crossings between two lowest states is the voltage at which extra charge is added or removed.

Figure 6 shows the corresponding approximation to the full charge stability diagram, the colours signifying the

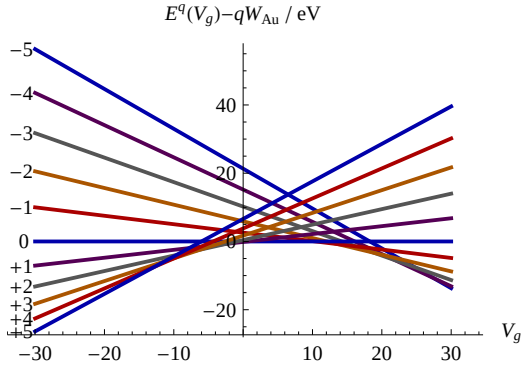


FIG. 5. Total energies corrected for the workfunction of gold. The lines are labeled with the charge q.

number of charge levels that are within the bias window $[-V_{sd}/2; V_{sd}/2]$. Within the black diamonds, the molecule is stable, and no charge can be added or removed; within the coloured regions, electrons flow to or from the electrodes. That is, the transistor is open around the crossings, where a source-drain voltage results in electron transport.

When $V_{sd} = 0$, the transition between charge state q and $q+1$ occur when the energy difference between the two states equal the work function. This is attained when

$$V_{q\to q+1} = -\frac{\Delta E^q(0) - W}{\alpha^q} \tag{1}$$

where $\alpha^q = \frac{1}{q}\frac{\partial E^q}{\partial V_g}$ is the gate coupling coefficient. The width of a charge state, i.e. the voltage between two transition points, is easily calculated to be

$$L_q \approx \frac{\Delta^2 E^{q-1}(0)}{\alpha^q} \tag{2}$$

given that $\alpha^{q+1} \approx \alpha^q$. The transition points, and thus the gate voltages at which conductance can occur when applying a source-drain potential, are thus determined entirely by the parameters $(\Delta E^q(0) - W)$ and α, and the widths of the charge states are controlled by $\Delta^2 E^q(0)$ and α.

1. *Effect of the SET dimensions*

Figures 7 and 8 show, how IP, EA, the addition energy $U = \Delta^2 E^{q-1}$, and the gate coupling coefficient, $\alpha = \frac{1}{q}\frac{\partial E}{\partial V_g}$, change with the distance d_y of the molecule to the gate oxide. In addition, the dependence of α on the gate thickness H is shown. The energies at $V_g = 0$ are largely unaffected by the gate thickness, so we omit this plot. α, on the other hand, is approximately inversely proportional to H.

As we saw from Equation (2), the width of the charge state q in the charge stability diagram is approximately $L_q \approx \Delta^2 E^{q-1}(0)/\alpha$. If we calculate the width L_0 of the

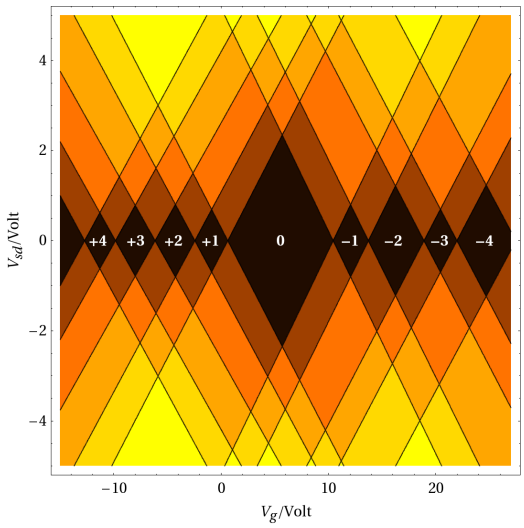


FIG. 6. Approximate charge stability diagram for OPV5-tBu in the SET-environment. The labels in the diamonds are added charge q; i.e., as the gate voltage V_g is increased, more electrons will move from the gold to the molecule.

neutral charge state, we can see that it grows approximately linearly both in d_y and in H:

$$\begin{aligned} L_0 &\approx 0.205\tfrac{V}{\text{Å}} \cdot H && \text{when } d_y = 1\text{Å}, H \geq 30\text{Å} \\ L_0 &\approx 8V + 2.27\tfrac{V}{\text{Å}} \cdot d_y && \text{when } H = 50\text{Å} \end{aligned} \tag{3}$$

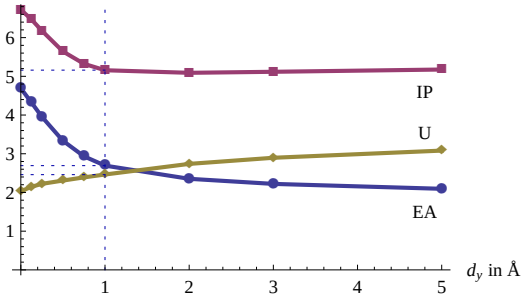


FIG. 7. The first ionization energy IP, elentron affinity EA, and addition energy U of OPV5-tBu as a function of the distance d_y to the gate oxide. The gate thickness is kept at a constant $H = 50$Å, and no external voltage is applied. The dotted lines show the values at $d_y = 1$Å, the distance used in our calculations.

VI. ANALYSIS OF THE RESULTS

2. *Understanding the energies in vacuum*

We can understand the charging and addition energies by way of the Kohn-Sham spectrum of the OPV5-tBu molecule. While, due to electron correlation and self interaction, Koopmans' theorem[17] cannot be directly applied to density functional theory, analogous results do

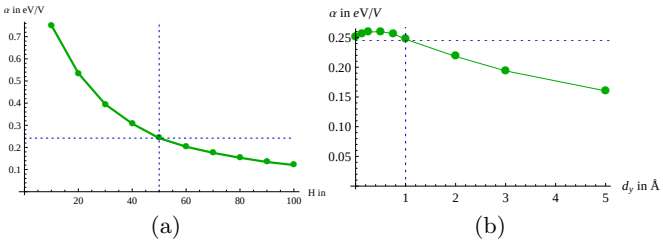


FIG. 8. The gate coupling coefficient $\alpha = \frac{1}{q}\frac{\partial E^{q+1}}{\partial V_g}$ of the OPV5-tBu molecule (a) as a function of gate thickness H at $d_y = 1$Å, and (b) the distance d_y from gate to molecule.

exist connecting the highest occupied Kohn-Sham eigenvalue to the IP[18–21]. Specifically, Perdew and Levy[19] showed that for the exact DFT solution, the vertical ionization energy $IP = -\varepsilon_N$; for the approximate DFT methods that are currently realizable, we have[21]

$$\Delta E = E^{N-1} - E^N \approx -\varepsilon_N^N + K[\rho^N] \tag{4}$$

where N is the total number of electrons, and the functional-dependent term K contains electrostatic and self-interaction terms. For the charge states of an extended molecule, K is nearly constant. For OPV5-tBu, we have found $K \approx -1eV$. The addition energies are then:

$$\Delta^2 E^N = \Delta E^{N-1} - \Delta E^N \approx \varepsilon_{N-1}^{N-1} - \varepsilon_N^N \tag{5}$$

Figure 9 shows Equation (5) to be in excellent agreement with the results.

We can further relate the charging and addition energies directly to the spectrum of Kohn-Sham eigenvalues of the *neutral* molecule in vacuum. The one-electron energies for the different charging states of OPV5-tBu are approximately linear in q with $\varepsilon_i^q \approx \varepsilon_i^0 - 2.1\text{eV}\cdot q$, whereby

$$\Delta E \approx -\varepsilon_i^q + K \approx -\varepsilon_i^0 + q \cdot 2.1\text{eV} - 1eV \tag{6}$$

and

$$\Delta^2 E \approx \varepsilon_{N-1}^0 - \varepsilon_N^0 + 2.1\text{eV} \tag{7}$$

That is, the addition energies $\Delta^2 E$ are well approximated simply by the one-electron spectrum shifted upwards by an additive constant corresponding to a quadratic contribution to the total energy due primarily to interelectron repulsion. We see from Figure 9 that Equation (7) reproduces the addition energies somewhat less accurately than Equation (5). However, the distinguishing features are preserved: the overall shape of $\Delta^2 E$ is closely matched to the spectrum of one-electron energies for the neutral molecule. As we will see below, quadratic energy contributions such as electrostatic effects merely shift $\Delta^2 E$ up or down by constant amounts, unless the dependence on q of the spatial distribution of charge is altered significantly, the molecular geometry changes, or another phenomenon with super-quadratic energy-contribution occurs.

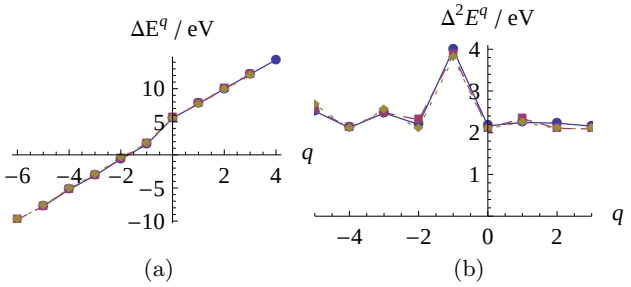


FIG. 9. Addition energies for the isolated OPV5-tBu molecule compared to approximations from Kohn-Sham eigenvalues: ——●——: ΔE^q_{Vac} (left) and $\Delta^2 E^q_{\text{Vac}}$ (right); - - ■ - -: Approximations (4) and (5); and ···◆···: Approximations (6) and (7).

3. *Understanding the energy shift from vacuum to SET*

Figure 3(b) shows that the shift in charging energies ΔE^q from vacuum to the SET environment is linear in q, i.e. the change in $\Delta^2 E^q$ is (mostly) constant. This can be explained using simple classical electrostatics: The molecule is stretched out, mostly flat, and has a large surface area lying closely against the gate dielectric; only the extreme points of the molecule are close to the source/drain-electrodes. At $V_{sd} = 0$, we consequently expect the dominant effect from the SET on the molecule to be the electrostatic interaction with the dielectric. We can then roughly approximate the shift by neglecting the effect of the electrodes and treating the problem with an induced image-charge distribution:

$$\begin{aligned} E^q_{\text{SET}} &\approx E^q_{\text{Vac}} + \frac{1}{2}\frac{\epsilon-1}{\epsilon+1}\iint \delta\rho(x_1)\frac{1}{r_{12}}\delta\rho_{Img}(x_2) \\ &= E^q_{\text{Vac}} - \frac{1}{2}\frac{\epsilon-1}{\epsilon+1}\iint \delta\rho(x_1)\frac{1}{r_{12}}\delta\rho(x_2-(0,0,2d_y)) \end{aligned}$$

where $\delta\rho \equiv \rho^q - \rho^0$. Assume now that the spatial distribution of added charge $\Delta\rho^q(x) = \rho^{q+1}(x) - \rho^q(x)$ is fixed, i.e. $\Delta\rho^{q+1} \approx \Delta\rho^q$. Then we can factor $-q^2$ out from the integral and obtain

$$\begin{aligned} E^q_{\text{SET}} &\approx E^q_{\text{Vac}} - q^2 C \\ \Delta E^q_{\text{SET}} &\approx \Delta E^q_{\text{Vac}} - (2q+1)C \\ \Delta^2 E^q_{\text{SET}} &\approx \Delta^2 E^q_{\text{Vac}} - 2C \end{aligned} \tag{8}$$

where the coefficient C is defined as $C = \frac{\epsilon-1}{2(\epsilon+1)} I(\Delta\rho, d_y)$, with I being the integral on the right hand side of Equation (VI 3). C depends only on the constant distribution of charge difference $\Delta\rho(x)$, the dielectric constant ϵ, and the distance d_y from molecule to dielectric. Note, that while we have only considered the contribution from the dielectric, if we were to include the electrodes or added other classical electrostatic element, it would not change Equation (8) qualitatively. So long that we assume a fixed $\Delta\rho$, only the coefficient C is affected. In this case, the contribution to the total energy must be $\propto q^2$, it must be $\propto (2q+1)$ to the charging energies ΔE^q, and constant to the addition energies $\Delta^2 E^q$.

We can give a quantitative estimate of the validity of our calculations in the SET environment with the following rough approximation: Assume a charge difference that is equally distributed over a two-dimensional rectangle A. Then

$$C(A) = \frac{9}{22|A|^2}\int_A \frac{1}{(x_1-x_2)^2 + (z_1-z_2)^2 + 4d_y^2} \tag{9}$$

and the shift in ionization energies is

$$\Delta E^q_{\text{SET}} - \Delta E^q_{\text{Vac}} \approx -(2q+1)C(A) \tag{10}$$

By setting the rectangle to the xz-bounding box of the OPV5-tBu molecule (38.24Å × 5.45Å), this approximation yields

$$\Delta E^q_{\text{SET}} - \Delta E^q_{\text{Vac}} \approx -0.740(2q+1)eV \tag{11}$$

which is to be compared to the value obtained from the full calculation

$$\Delta E^q_{\text{SET}} - \Delta E^q_{\text{Vac}} = -0.738(2q+0.528) \tag{12}$$

The small difference in the coefficient C is in part due to the crude approximation of assuming a uniform rectangular charge distribution, in part due to inclusion of the entire SET-environment; However, as is apparent from the discussion above, the difference in the intersection with the q-axis ($2q = -0.53$ as opposed to $2q = -1$) must be due to charge localization that varies with q, since otherwise the factor $(2q+1)$ must appear. Similar calculations for the more compact benzene molecule yields a crossing at $2q = -0.8$, and for the hydrogen ions close to $2q = -1$, substantiating this claim.

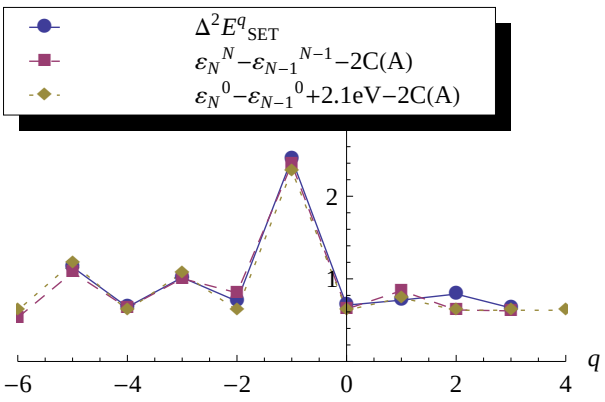


FIG. 10. Addition energies for OPV5-tBu in the SET environment compared to approximations calculated from Equation (9) and Equation (5) and (7), respectively.

VII. COMPARING TO THE EXPERIMENTS OF KUBATKIN ET AL.

Kubatkin et al.[22] succeeded in constructing a molecular SET using a single OPV5-tBu molecule as the capacitor.

A. Reported results

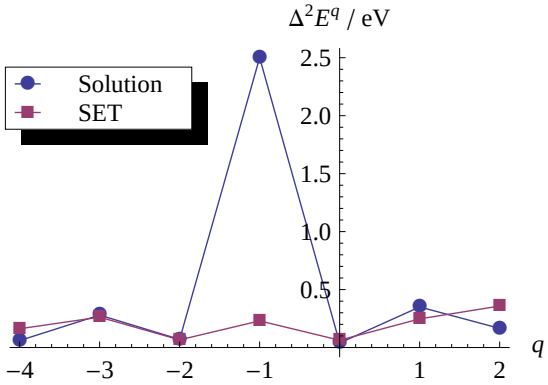


FIG. 11. Addition energies reported by Kubatkin et al.[22] The values in the solution are comprised of estimates from semiempirical AM1 calculations (right of peak), values obtained from electrochemical midway potentials[23] (left of peak), and estimate from the absorption edge of OPV5 in the solution. The SET-values are all experimental and derived from the charge stability diagram of the constructed single electron transistor.

There are two remarkable features of the numbers reported by Kubatkin et al., as we can see on Figure 11:

i. The peak representing the addition energy $\Delta^2 E^{-1}$ = IP − EA has disappeared in the SET environment.

ii. Both measurements and calculations in the solution, and the measurements in the SET environment yield very small values of $\Delta^2 E^q$.

Why are these two features so remarkable? Ad (i), we recall from Section VI 3 that any energy contribution that is up to quadratic in q (such as simple electrostatic effects) must only shift $\Delta^2 E^q$ up or down by a constant amount. To reduce the IP-EA gap $\Delta^2 E^{-1}$ to almost zero, while at the same time leaving all the other addition energies essentially unchanged, requires a large superquadratic and likely non-polynomial effect. Ad (ii), the small values correspond to total energy functions with very low curvature. In the solution, the numbers predict a total energy curve that is nearly piecewise linear in the given region, with a sharp break at $q = 0$. However, for the SET environment, the numers yield a total energy curve that is nearly linear everywhere! If the addition energies measured by Kubatkin et al. are correct, then both of these are very interesting results that suggest complex effects at play. In the next subsection, we will see the dramatic effect of Point (ii) on the autoionization properties of the OPV5-tBu molecule.

B. Autoionization

Due to the reduction in addition energies in the electrostatic environment, the autoionization properties of the OPV5-tBu molecule are altered when the molecule is deposited in the SET even when no external voltages are applied. Reducing the addition energies corresponds to reducing the curvature of the total energy as a function of q, as was seen on Figure 4. Consequently, the energy minimum changes position. The minimum of the total energy curve, however, does not directly yield the spontaneous level of ionization: We must take into account the energy needed to remove an electron from the environment. The most stable charge state has charge

$$q_0 = \operatorname*{argmin}_q \left(E^q - qW\right) \tag{13}$$

where W is the work function for the environment surrounding the molecule. That is, the molecule will spontaneously release or obtain electrons until it reaches the charge state q_0.

In the SET environment, we can assume that electrons will come from the gold electrodes and let W be the work function for gold; A good value is $\mathsf{W}_{\mathrm{Au}} = 5.28\mathrm{eV}$ recommended by Rivière.[16] However, depending on temperature, purity, surface properties and multiple other factors, the work function for gold can vary up to a few electron volts. In order to predict the range of autoionization one will find in experiments, it is then necessary to allow for a variance in W. To assess the level of autoionization found in Kubatkin et al.[22], we further need to allow for a variance in the ionization energy: Whatever causes the large difference in addition energies for the Kubatkin et al. setup may also cause the ionization energy to be quite changed. Since no charging energies were reported, we must leave IP as a free parameter in the following.

Before we continue, we note that we can reconstruct the charging energies and total energy curves from addition energies using the relation

$$\sum_{q=a}^{b-1} \Delta f_q = f_b - f_a \tag{14}$$

That is, given a single value ΔE^{q_0}, we can derive all other values ΔE^q from the $\Delta^2 E^q$'s, et cetera. Applying Equation (14) twice yields:

$$E^q = E^0 + q\Delta E^0 + \sum_{n=0}^{q-1} \sum_{m=0}^{n-1} \Delta^2 E^m \tag{15}$$

In other words, we can write the work function corrected energy as

$$E^q - qW = E^0 + \sum_{n=0}^{q-1}\sum_{m=0}^{n-1} \Delta^2 E^m + q(\mathrm{IP} - W) \qquad (16)$$

Consequently, we do not need to vary W and the ionization energy separately; Varying the parameter IP $- W$ suffices.

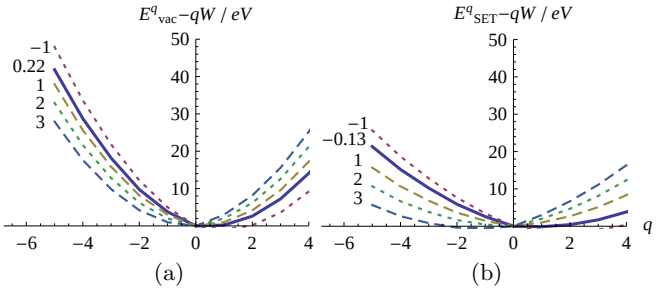


FIG. 12. Total energy curve with work function correction (a) in vacuum, and (b) in the SET. The energies were derived from the calculated $\Delta^2 E^q$ given the specified values of $IP - W$ (given in eV). The minima of the curves, which are listed in Table IV, predict the level of autoionization. The solid curve corresponds to our calculated IP and $W = 5.28$eV.

Figure 12 shows the work function corrected energy curves corresponding to the range $IP - W \in [-1\mathrm{eV}; 3\mathrm{eV}]$ for our FEM+LDA calculations. The minima, corresponding to the autoionized charge states, are given in Table IV. We observe that while the lowered addition energies in the SET environment do make the molecule more volatile, it is mostly stable both in vacuum and deposited on the SET. Within the most common range of W_{Au}, it is predicted to either remain in its neutral form or donate a single electron. Assuming the calculated IP is correct, a work function of around 2.5eV for the environment is required for the OPV5-tBu to spontaneously acquire an electron, while the work function must be around 6eV for the molecule to donate a single electron.

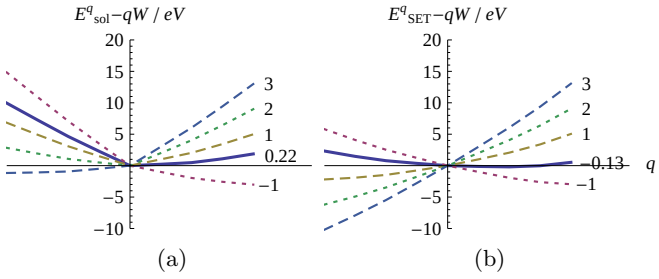


FIG. 13. Total energies with work function correction derived from the addition energies reported by Kubatkin et al. (a) reported for OPV5-tBu in a solution, and (b) measured for OPV5-tBu in the SET environment. The values for $IP - W$ are the same as above. Note the change of scale compared to Figure 12.

Calculated with FEM+LDA							
IP$-W$ in eV		-1	0	1	2	3	[a]
Stable charge state q_0	in vacuum	+1	0	0	0	0	0
	in SET	+2	0	0	0	-1	+1

Kubatkin et al.							
IP$-W$ in eV		-1	0	1	2	3	[a]
Stable charge state q_0	in solution	+5	0	0	0	-4	0
	in SET	+6	0	-5	-10	-16	+2

[a] Calculated IP minus $\mathsf{W}_{\mathrm{Au}} = 5.28$eV: $5.55\mathrm{eV} - 5.28\mathrm{eV} = 0.22$eV in vacuum, and $5.15\mathrm{eV} - 5.28\mathrm{eV} = -0.13$eV in the SET.

TABLE IV. Autoionization of the OPV5-tBu molecule derived from addition energies for a range of IP $- W$, where W is the work function for the environment. The lower table is derived for from the addition energies reported by Kubatkin et al.[22]. For charges lower than $q = -4$, values for $\Delta^2 E^q$ were not available; In this case, a quadratic polynomial was fitted to the total energies for $q \leq 0$, the minimum of which predicts the level of ionization.

The situation in the experimental setup is in quite strong contrast to this: As previously noted, the shallow addition energies measured and calculated by Kubatkin et al. yield a very low curvature of the total energy as a function of q. Consequently, small changes in the absolute value of the work function or ionization energy lead to drastically different autoionization properties. This is especially true in the SET environment, where there is no peak at $\Delta^2 E^{-1} = IP - EA$. As can be seen from Figure 13 and Table IV, the measured addition energies predict the OPV5-tBu molecule to be moderately stable in the solution, but wildly volatile when placed in the SET environment: The numbers predict that the molecule spontaneously gives up 6 electrons at IP $- W = -1$eV, but acquires 5 electrons at IP $- W = 1$eV, a change of 11 electrons due to a difference in IP $- W$ of just 2eV, a variation that is not unreasonable for experimental setups. If the measured addition energies are correct, it is thus possible that a completely different part of the spectrum of OPV5-tBu was measured than that around the neutral charge state. To determine *which* part of the spectrum this is, it is necessary to determine actual ionization energy of OPV5-tBu in the experimental SET environment, as well as the work function.

C. Discussion

In summary, there are a number of possible reasons that calculations did not reproduce the effect observed by Kubatkin et al.

1. Theoretically, it is possible that strong charge localization can cause the highly nonlinear reduction in addition energies. However, for electrostatic effects to almost exactly cancel the first addition energy $U = \Delta^2 E^{-1}$ while at the same time keeping all the other addition energies constant, quite elabo-

rate conditions must hold. We have attempted a number of rough calculations taking such an effect into account, and have found it very difficult to reproduce anything close to this behaviour.

2. As the previous section showed, the extremely low addition energies measured by Kubatkin et al. correspond to a very volatile molecule that easily gains or relinquishes many electrons. The missing peak for the first addition energy may then be due to having observed a separate part of the spectrum.

3. It is possible that the environment in the experimental setup introduce a change of molecular conformation, causing the drastic change of spectrum. Our calculations use a fixed geometry optimized for OPV5-tBu in vacuum, and thus would not reproduce such an effect.

VIII. CONCLUSION

* avery@diku.dk
† stig@diku.dk
‡ kaasbjerg@dtu.dk
§ m.lein@massey.ac.nz
¶ kurt@quantumwise.com

[1] K. Moth-Poulsen and T. Bjørnholm, Nature Nanotech. **4**, 551 (2009).
[2] J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruña, P. L. McEuen, and D. C. Ralph, Nature **417**, 722 (2002).
[3] S. Kubatkin, A. Danilov, M. Hjort, J. Cornil, J.-L. Brédas, N. S.-Hansen, P. Hedegård, and T. Bjørnholm, Nature **425**, 698 (2003).
[4] L. H. Yu, Z. K. Keane, J. W. Ciszek, L. Cheng, J. M. Tour, T. Baruah, M. R. Pederson, and D. Natelson, Phys. Rev. Lett. **95**, 256803 (2005).
[5] E. A. Osorio, K. O'Neill, M. Wegewijs, N. Stuhr-Hansen, J. Paaske, T. Bjørnholm, and H. S. J. van der Zant, Nano. Lett. **7**, 3336 (2007).
[6] E. A. Osorio, K. O'Neill, N. Stuhr-Hansen, O. F. Nielsen, T. Bjørnholm, and H. S. J. van der Zant, Adv. Mater. **19**, 281 (2007).
[7] K. Kaasbjerg and K. Flensberg, Nano. Lett. **8**, 3809 (2008).
[8] S. S. Datta, D. R. Strachan, and A. T. C. Johnson, Phys. Rev. B **79**, 205404 (2009).
[9] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, Journal of Physics: Condensed Matter **14**, 2745 (2002).
[10] K. Kaasbjerg and K. Flensberg, Nano Letters **8**, 3809 (2008), dOI: 10.1021/nl8021708.
[11] J. Avery, "A linearly scaling finite element based dft," In progress. (2011).
[12] OPV5 ($C_{46}H_{46}S_2$): ((E,E)-1,4-bis{4-{(E)-4-(tertbutylthio)styryl}}benzene) with added tert-butyl groups protecting the terminal thiols and preventing the sulphur from binding to the gold electrodes.
[13] M. J. Frisch, G. W. Trucks, H. B. Schlegel, and G. E. S. et al., "Gaussian 03, Revision B.05," (2003), Gaussian, Inc., Pittsburg, PA, 2003.
[14] V. Papaefthimiou, S. Siokou, and S. Kennou, Journal of Applied Physics **91**, 4213 (2002).
[15] NIST CCCBDB, "Nist computational chemistry comparison and benchmark database: Nist standard reference database number 101," http://cccbdb.nist.gov/ (2010), release 15a.
[16] J. C. Rivière, Applied Physics Letters **8**, 172 (1966).
[17] T. Koopmans, Physica **1**, 104 (1934).
[18] J. F. Janak, Physical Review B **18**, 7166 (1978).
[19] J. P. Perdew and M. Levy, Physical Review B **56**, 16021 (1997).
[20] P. Politzer and F. Abu-Awwad, Theoretical Chemistry Accounts **99**, 83 (1998).
[21] J. Luo, Z. Q. Xue, W. M. Lui, J. L. Wu, and Z. Q. Yang, Journal of Physical Chemistry A **2006**, 12005 (2006).
[22] S. Kubatkin, A. Danilov, M. Hjort, J. Cornil, J.-L. Bredas, N. Stuhr-Hansen, P. Hedegard, and T. Bjornholm, Nature **425**, 698 (2003).
[23] J. Heinze, J. Mortensen, K. Müllen, and R. Schenk, J. Chem. Soc. Chem. Commun. , 701 (1987).

Publications on Sturmians

Solving the Schrödinger Equation: Has everything been tried? Chapter 6: The generalized Sturmian method

Solving the Schrödinger Equation: Has everything been tried? (Popelier [2011]), brings together 10 authors of non-standard but promising methods in quantum chemistry.

> The Schrödinger equation is the master equation of quantum chemistry. The founders of quantum mechanics realised how this equation underpins essentially the whole of chemistry. However, they recognised that its exact application was much too complicated to be soluble at the time. More than two generations of researchers were left to work out how to achieve this ambitious goal for molecular systems of ever-increasing size. This book focuses on non-mainstream methods to solve the molecular electronic Schrödinger equation. Each method is based on a set of core ideas and this volume aims to explain these ideas clearly so that they become more accessible. By bringing together these non-standard methods, the book intends to inspire graduate students, postdoctoral researchers and academics to think of novel approaches. Is there a method out there that we have not thought of yet? Can we design a new method that combines the best of all worlds?

Chapter 6

The Generalized Sturmian Method

James Avery and John Avery

The generalized Sturmian method makes use of basis sets that are solutions to an approximate wave equation with a weighted potential. The weighting factors are chosen in such a way as to make all the members of the basis set isoenergetic. In this chapter we will show that when the approximate potential is taken to be that due to the attraction of the bare nucleus, the generalized Sturmian method is especially well suited for the calculation of large numbers of excited states of few-electron atoms and ions. Using the method we shall derive simple closed-form expressions that approximate the excited state energies of ions. The approximation improves with increasing nuclear charge. The method also allows automatic generation of near-optimal symmetry adapted basis sets, and it avoids the Hartree-Fock SCF approximation. Programs implementing the method may be freely downloaded from our website, `sturmian.kvante.org` [Avery and Avery (2006)].

6.1 Description of the method

6.1.1 *The introduction of Sturmians into quantum theory*

One of the very early triumphs of quantum theory was the exact solution of the Schrödinger equation for hydrogenlike atoms:

$$\left[-\frac{1}{2}\nabla^2 - \frac{Z}{r} - E_n\right]\psi_{n,l,m}(\mathbf{x}) = 0 \tag{6.1}$$

In Equation (6.1) and throughout the chapter, atomic units are used. The energies and wavefunctions are given respectively by

$$E_n = -\frac{Z^2}{2n^2}, \qquad n = 1, 2, 3, \ldots, \tag{6.2}$$

and

$$\psi_{n,l,m}(\mathbf{x}) = R_{n,l}(r) Y_{l,m}(\theta, \phi) \tag{6.3}$$

Here $Y_{l,m}(\theta, \phi)$ is a spherical harmonic, and

$$\begin{aligned} R_{1,0}(r) &= 2(Z/1)^{3/2} e^{-Zr/1} \\ R_{2,0}(r) &= 2(Z/2)^{3/2}(1 - Zr/2) e^{-Zr/2} \\ R_{2,1}(r) &= \frac{2}{\sqrt{3}}(Z/2)^{3/2}\ (Zr/2)\ e^{-Zr/2} \\ &\vdots \quad \vdots \quad \vdots \end{aligned} \tag{6.4}$$

It was natural to try to use hydrogenlike orbitals as building blocks to represent the wave functions of more complicated atoms. However, to the great disappointment of the early workers in atomic theory, it was soon realized that unless the continuum was included the hydrogenlike orbitals did not form a complete set; and the continuum proved to be prohibitively difficult to use in practical calculations. This dilemma led Høloien, Shull and Löwdin [Shull and Löwdin (1959)] to introduce basis functions that have exactly the same form as hydrogenlike orbitals except that Z/n is replaced by a constant, k, which is the same for all the members of the basis set. This type of basis set came to be called *Coulomb Sturmians*, the name being given to them by A. Rotenberg [Rotenberg (1970)] to emphasize their connection with the Sturm-Liouville theory of orthonormal sets of functions. Coulomb Sturmian basis sets are complete *without the inclusion of the continuum*: any square-integrable solution to a one-electron Schrödinger equation can be represented as a linear superposition of them. If the potential in the one-electron Schrödinger equation has some similarity to a Coulomb potential - for example if it is a screened Coulomb potential - the convergence of such a series is rapid.

The members of a Coulomb Sturmian basis set are solutions to a one-electron equation of the form

$$\left[-\frac{1}{2}\nabla^2 - \frac{nk}{r} + \frac{k^2}{2}\right] \chi_{n,l,m}(\mathbf{x}) = 0 \tag{6.5}$$

If we compare Equation (6.5) with (6.1) we can see that with the substitutions $Z/n \to k$ and $E_n \to -k^2/2$, Equation (6.1) is converted into equation

(6.5). Therefore, if we interpret $-k^2/2$ as the energy, the solutions to the Coulomb Sturmian wave equation (6.5) are just the familiar hydrogenlike orbitals with Z/n replaced by k. In other words, they have the form

$$\chi_{n,l,m}(\mathbf{x}) = R_{n,l}(r)Y_{l,m}(\theta,\phi) \tag{6.6}$$

where Z/n in Equation (6.4) is replaced by a constant k, which is the same for all the members of the basis set. Since $-k^2/2$ is interpreted as the energy, all the members of a Coulomb Sturmian basis set correspond to the same energy: they are *isoenergetic*. The first few Coulomb Sturmian wave functions are shown in Table 6.1. You can easily see that if you make the substitution $k \to Z/n$ for the radial functions in this table, you will just get the familiar hydrogenlike atomic orbitals; but Coulomb Sturmian basis sets have very different properties! They obey a potential-weighted orthonormality relation:

$$\int d^3x\ \chi^*_{\mu'}(\mathbf{x})\frac{1}{r}\chi_{\mu}(\mathbf{x}) = \frac{k}{n}\delta_{\mu',\mu} \qquad \mu \equiv (n,l,m) \tag{6.7}$$

from which it follows that

$$\int d^3x\ \chi^*_{\mu'}(\mathbf{x})\left(\frac{-\nabla^2+k^2}{2k^2}\right)\chi_{\mu}(\mathbf{x}) = \delta_{\mu',\mu} \tag{6.8}$$

The Coulomb Sturmian basis sets behave quite differently from the usual sets of eigenfunctions to the zeroth-order Hamiltonian often used in quantum theory. Equation (6.1) is the usual type of eigenvalue problem with which everyone in the physical sciences is familiar. By contrast, Equation (6.5) is an entirely different problem, sometimes called a *conjugate eigenvalue problem*: Each member of a set of solutions corresponds to the same energy $-k^2/2$, k being a constant that is the same for all the members of the set. The quantity that plays the role of the usual eigenvalue is now a weighting factor attached to the potential, which is chosen in such a way as to make all the members of the basis set isoenergetic. Because of their useful properties, Coulomb Sturmian basis sets are widely used in atomic theory, and there exists a large literature discussing their properties and applications ([Shull and Löwdin (1959)] through [Sherstyuk (1983)], [Aquilanti *et al.* (1996a)] through [Avery and Herschbach (1992)], [Avery (2003)], [Caligiana (2003)], [Koga and Matsuhashi (1987)], and [Koga and Matsuhashi (1988)]).

Table 6.1: One-electron Coulomb Sturmian radial functions. If k is replaced by Z/n they are identical to the familiar hydrogenlike radial wave functions.

n	l	$R_{n,l}(r)$
1	0	$2k^{3/2}e^{-kr}$
2	0	$2k^{3/2}(1-kr)e^{-kr}$
2	1	$\frac{2k^{3/2}}{\sqrt{3}}\ kr\ e^{-kr}$
3	0	$2k^{3/2}\left(1-2kr+\frac{2(kr)^2}{3}\right)e^{-kr}$
3	1	$2k^{3/2}\frac{2\sqrt{2}}{3}\ kr\ \left(1-\frac{kr}{2}\right)e^{-kr}$
3	2	$2k^{3/2}\frac{\sqrt{2}}{3\sqrt{5}}\ (kr)^2\ e^{-kr}$

6.1.2 *Generalized Sturmians*

In 1968, Osvaldo Goscinski generalized the Sturmian concept by introducing basis sets that are solutions to an approximate many-particle Schrödinger equation with a weighted potential:

$$\left[-\sum_{j=1}^{N}\frac{1}{2m_j}\nabla_j^2+\beta_\nu \mathsf{V}_0(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N)-E\right]\Phi_\nu(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N)=0 \quad (6.9)$$

the weighting factor β_ν once again being chosen in such a way as to make all of the solutions correspond to the same energy. When $N = 1$, $\mathsf{V}_0(\mathbf{x}) = -Z/r$, and $\beta_\nu = nk/Z$, this equation reduces to Equation (6.5), obeyed by Coulomb Sturmians. Basis sets of this kind have many advantages, especially the advantage of spanning an appropriate Hilbert space, and they could potentially be used in a wide variety of problems; but until now the applications of generalized Sturmians have been very limited because most physicists and theoretical chemists are unfamiliar with them. In fact Osvaldo Goscinski did not publish his pioneering 1968 paper until very recently [Goscinski (2002)]. It was only printed as an internal report of the Uppsala University Quantum Chemistry Group, and was known to just a small circle of people. The idea remained dormant, and unfortunately little use was made of generalized Sturmian basis sets in practical calculations. Today, however, the generalized Sturmian method is an idea whose time has come!

In the generalized Sturmian method, the basis functions are chosen to be isoenergetic solutions to an approximate Schrödinger equation with a weighted potential (references [Aquilanti *et al.* (1996b)]-[Avery and Avery (2007)]).

A set of generalized Sturmian basis functions can be shown ([Goscinski (2002)], or [Avery and Avery (2007)], Chapter 1) to obey the following potential-weighted orthonormality relations:

$$\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_\nu(\mathbf{x}) = \delta_{\nu'\nu}\frac{2E_\kappa}{\beta_\nu} = -\delta_{\nu'\nu}\frac{p_\kappa^2}{\beta_\nu} \quad \text{where} \quad p_\kappa^2 \equiv -2E_\kappa \tag{6.10}$$

Where we let κ denote a particular state and where we have introduced the abbreviated notation $\mathbf{x} \equiv (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$. To obtain the generalized Sturmian secular equations, we begin by substituting the superposition

$$\Psi_\kappa(\mathbf{x}) = \sum_\nu \Phi_\nu(\mathbf{x})B_{\nu\kappa} \tag{6.11}$$

into the Schrödinger equation (6.22). This yields:

$$\sum_\nu \left[-\sum_{j=1}^{N} \frac{1}{2m_j}\nabla_j^2 + \mathsf{V}(\mathbf{x}) - E_\kappa \right] \Phi_\nu(\mathbf{x})B_{\nu\kappa} = 0 \tag{6.12}$$

We now split the potential $\mathsf{V}(\mathbf{x})$ into two parts, $\mathsf{V}(\mathbf{x}) = \mathsf{V}_0(\mathbf{x}) + \mathsf{V}'(\mathbf{x})$, and introduce the definitions

$$\begin{aligned} T^0_{\nu'\nu} &\equiv -\frac{1}{p_\kappa} \int d\tau\ \Phi^*_{\nu'}(\mathbf{x}) \mathsf{V}_0(\mathbf{x}) \Phi_\nu(\mathbf{x}) \\ T'_{\nu'\nu} &\equiv -\frac{1}{p_\kappa} \int d\tau\ \Phi^*_{\nu'}(\mathbf{x}) \mathsf{V}'(\mathbf{x}) \Phi_\nu(\mathbf{x}) \end{aligned} \tag{6.13}$$

From the potential-weighted orthonormality relations (6.10) it follows that T^0 is diagonal:

$$T^0_{\nu'\nu} = \delta_{\nu'\nu} T^0_{\nu\nu} = \delta_{\nu'\nu} \frac{p_\kappa}{\beta_\nu} \tag{6.14}$$

Next, we notice that since all of the isoenergetic configurations in the basis set obey (6.9), Equation (6.12) can be rewritten as

$$\sum_\nu \left[\mathsf{V}(\mathbf{x}) - \beta_\nu \mathsf{V}_0(\mathbf{x})\right] \Phi_\nu(\mathbf{x}) B_{\nu\kappa} = 0 \tag{6.15}$$

We then multiply by a conjugate function from our basis set and integrate over all space and spin coordinates:

$$\sum_\nu \int d\tau\ \Phi^*_{\nu'}(\mathbf{x}) \left[\mathsf{V}(\mathbf{x}) - \beta_\nu \mathsf{V}_0(\mathbf{x})\right] \Phi_\nu(\mathbf{x}) B_{\nu\kappa} = 0 \tag{6.16}$$

Making use of Equations (6.13)-(6.14), we obtain

$$\sum_\nu \left[-p_\kappa T^0_{\nu'\nu} - p_\kappa T'_{\nu'\nu} + \beta_\nu p_\kappa T^0_{\nu'\nu}\right] B_{\nu\kappa} = 0 \tag{6.17}$$

Using (6.14) to derive

$$\beta_\nu p_\kappa T^0_{\nu'\nu} = \delta_{\nu'\nu} p^2_\kappa \tag{6.18}$$

and finally, dividing by $-p_\kappa$, we obtain the generalized Sturmian eigenproblem:

$$\sum_\nu \left[\delta_{\nu'\nu} T^0_{\nu\nu} + T'_{\nu'\nu} - p_\kappa \delta_{\nu'\nu}\right] B_{\nu\kappa} = 0 \tag{6.19}$$

Generalized Sturmian basis sets can come in many species and varieties: Every choice of the approximate potential $\mathsf{V}_0(\mathbf{x})$ (which should be chosen to resemble $\mathsf{V}(\mathbf{x})$ as closely as possible) leads to a particular set of *shapes* for the N-particle basis functions $\Phi_\nu(\mathbf{x})$. Solving Equation (6.9), which is

done once and for all for a particular V_0, specifies the functions Φ_ν up to an undetermined scaling parameter p_κ. Solving the generalized Sturmian eigenproblem (6.19) then yields as eigenvalues the scaling parameters p_κ and as eigenfunctions

$$\Psi_\kappa(\mathbf{x}) = \sum_\nu B_{\nu\kappa}\Phi_\nu(p_\kappa, \mathbf{x}) \tag{6.20}$$

where each p_κ scales the entire basis to give all the N-particle basis functions the same energy E_κ. If the generalized Sturmian basis $\{\Phi_\nu\}$ is complete, then Equation (6.19) has exactly the same eigenfunctions Ψ_κ as the Schrödinger equation, and the energies are

$$E_\kappa = -\frac{p_\kappa^2}{2} \tag{6.21}$$

In practice, one of course always uses a finite basis, so solutions are approximate. However, we shall see that the automatic scaling allows us to obtain good accuracy with few basis functions, as well as to obtain many excited states at once.

It is remarkable to see how completely Equation (6.19) differs from the conventional secular equations used in quantum theory:

(1) The kinetic energy term has vanished.
(2) The matrix representing the approximate potential $\mathsf{V}_0(\mathbf{x})$ is diagonal.
(3) The roots are not energies but values of the *scaling parameter*, p_κ, which is proportional to the square roots of the binding energies (Equation (6.10)).
(4) Before the secular equation is solved, only the shapes of the basis functions are known, but not the values of the scaling parameters p_κ.
(5) Solution of the secular equations yields a near-optimum basis set appropriate for each state, as well as the states themselves and their corresponding energies.
(6) The Hamiltonian formalism is nowhere to be seen!

In the present chapter, we review the generalized Sturmian method applied to atoms and atomic ions, as well as the large-Z approximation that

was introduced by us in [Avery and Avery (2005)]. These methods have been described in much more detail in our recent book [Avery and Avery (2007)]. Other work on Sturmians and generalized Sturmians can be found in references [Aquilanti *et al.* (1996b)]-[Koga and Matsuhashi (1988)].

The large-Z approximation yields extremely simple closed form expressions for the approximate energies of both the ground states and excited states of atoms and atomic ions. The accuracy of the large-Z approximation for few-electron systems is such that even for moderate values of Z, inaccuracies are much smaller than relativistic corrections. An approximate method for making relativistic corrections is introduced below in Section 6.2.2. It is shown that the corrected energies rapidly approach the experimental ones as Z increases.

6.1.3 *The generalized Sturmian method applied to atoms*

In atomic units, the non-relativistic Schrödinger equation for an N-electron atom or atomic ion with nuclear charge Z is given by

$$\left[-\frac{1}{2}\Delta + \mathsf{V}(\mathbf{x}) - E_\kappa\right]\Psi_\kappa(\mathbf{x}) = 0 \tag{6.22}$$

where Ψ_κ is the κ'th electronic state with E_κ the corresponding energy, and where $\mathbf{x}$ stands for all the coordinates including spin.

$$\mathsf{V}(\mathbf{x}) = -\sum_{j=1}^{N}\frac{Z}{r_j} + \sum_{j>i}^{N}\sum_{i=1}^{N}\frac{1}{r_{ij}} \tag{6.23}$$

and

$$-\frac{1}{2}\Delta \equiv -\frac{1}{2}\sum_{j=1}^{N}\nabla_j^2 \tag{6.24}$$

where j is the index of an individual electron in the system.

6.1.4 *Goscinskian configurations*

When the generalized Sturmian method is applied to atoms or atomic ions, a wonderful thing happens: We wish to use a basis set consisting of solutions to

$$\left[-\frac{1}{2}\Delta + \beta_\nu \mathsf{V}_0(\mathbf{x}) - E_\kappa\right]\Phi_\nu(\mathbf{x}) = 0 \qquad E_\kappa = -\frac{p_\kappa^2}{2} \tag{6.25}$$

It turns out that if we choose $\mathsf{V}_0(\mathbf{x})$ to be the Coulomb attraction potential of the nucleus, then exact solutions to Equation (6.25) can be found with the greatest ease! Furthermore, the weighting factors β_ν are obtained automatically. And as if this were not enough, there is a final bonus: The basis functions $\Phi_\nu(\mathbf{x})$ are automatically normalized! How can all this be possible? Read on and see.

As just mentioned, we let $\mathsf{V}_0(\mathbf{x})$ be the electrostatic attraction potential of the nucleus:

$$\mathsf{V}_0(\mathbf{x}) = -\sum_{j=1}^{N} \frac{Z}{r_j} \quad \text{and} \quad \mathsf{V}'(\mathbf{x}) = \sum_{j>i}^{N} \sum_{i=1}^{N} \frac{1}{r_{ij}} \tag{6.26}$$

Now we claim that with this choice of $\mathsf{V}_0(\mathbf{x})$, the weighting factors β_ν are determined automatically, and Equation (6.25) is satisfied by Slater determinants of the form:

$$\Phi_\nu(\mathbf{x}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_{\mu_1}(1) & \chi_{\mu_2}(1) & \cdots & \chi_{\mu_N}(1) \\ \chi_{\mu_1}(2) & \chi_{\mu_2}(2) & \cdots & \chi_{\mu_N}(2) \\ \vdots & \vdots & & \vdots \\ \chi_{\mu_1}(N) & \chi_{\mu_2}(N) & \cdots & \chi_{\mu_N}(N) \end{vmatrix} = |\chi_{\mu_1}\chi_{\mu_2}\cdots\chi_{\mu_N}| \tag{6.27}$$

where the χ_μ's are just the familiar hydrogenlike spin-orbitals

$$\begin{aligned} \chi_{nlm,+1/2}(\mathbf{x}_j) &= R_{nl}(r_j)Y_{lm}(\theta_j,\phi_j)\alpha(j) \\ \chi_{nlm,-1/2}(\mathbf{x}_j) &= R_{nl}(r_j)Y_{lm}(\theta_j,\phi_j)\beta(j) \end{aligned} \tag{6.28}$$

but with the weighted charges Q_ν (Reference [Avery and Avery (2007)], Chapter 3) chosen according to the rules in the following box, where $n_1, n_2, \ldots, n_N$ are the principal quantum numbers of the hydrogenlike spin-orbitals in the configuration Φ_ν. The Goscinskian configurations will be *exact* solutions to (6.25) provided that:

$$\begin{aligned} Q_\nu &= \beta_\nu Z = \frac{p_\kappa}{\mathcal{R}_\nu} \\ p_\kappa &\equiv \sqrt{-2E_\kappa} \\ \mathcal{R}_\nu &\equiv \sqrt{\frac{1}{n_1^2} + \frac{1}{n_2^2} + \cdots + \frac{1}{n_N^2}} \end{aligned} \tag{6.29}$$

At this point the reader may be muttering "I don't believe it". Well, if you don't believe it, think about this: The energy E_κ will then be related to the weighted nuclear charges Q_ν by

$$E_\kappa = -\frac{p_\kappa^2}{2} = -\frac{1}{2}Q_\nu^2 {\mathcal{R}_\nu}^2 = -\left(\frac{Q_\nu^2}{2n_1^2} + \frac{Q_\nu^2}{2n_2^2} + \cdots + \frac{Q_\nu^2}{2n_N^2}\right) \tag{6.30}$$

Each of the hydrogenlike spin-orbitals obeys a one-electron Schrödinger equation of the form:

$$\left[-\frac{1}{2}\nabla_j^2 + \frac{Q_\nu^2}{2n^2} - \frac{Q_\nu}{r_j}\right]\chi_\mu(\mathbf{x}_j) = 0 \tag{6.31}$$

From Equation (6.31) it follows that

$$\begin{aligned}
&\left[-\frac{1}{2}\sum_{j=1}^{N}\nabla_j^2\right]\Phi_\nu(\mathbf{x}) \\
&= \left[-\left(\frac{Q_\nu^2}{2n_1^2} + \frac{Q_\nu^2}{2n_2^2} + \cdots + \frac{Q_\nu^2}{2n_N^2}\right) + \left(\frac{Q_\nu}{r_1} + \frac{Q_\nu}{r_2} + \cdots + \frac{Q_\nu}{r_N}\right)\right]\Phi_\nu(\mathbf{x}) \\
&= \left[E_\kappa - \beta_\nu \mathsf{V}_0(\mathbf{x})\right]\Phi_\nu(\mathbf{x})
\end{aligned} \tag{6.32}$$

Now compare Equation (6.32) with (6.25). They are the same! Thus Equation (6.25) will indeed be satisfied by the configurations Φ_ν shown in Equation (6.27), provided that the effective nuclear charges Q_ν are chosen according to the rule given in Equation (6.29). We shall call such a set of isoenergetic solutions to (6.25) – with $\mathsf{V}_0(\mathbf{x})$ chosen to be the nuclear attraction potential – a set of "*Goscinskian configurations*" to honor Prof. Osvaldo Goscinski's important early contributions to the generalized Sturmian method [Goscinski (2002)].

6.1.5 *Goscinskian secular equations for atoms and atomic ions*

Recall the solution of β_ν from Equation (6.29)

$$\beta_\nu = \frac{p_\kappa}{Z\mathcal{R}_\nu} \quad \text{and} \quad p_\kappa = \sqrt{-2E_\kappa} \tag{6.33}$$

This, and the potential-weighted orthonormality relations (6.10), give

$$T^0_{\nu'\nu} = \delta_{\nu'\nu} Z\mathcal{R}_\nu \tag{6.34}$$

Thus the matrix $T^0_{\nu'\nu}$ is diagonal and independent of p_κ. It can be shown ([Avery and Avery (2007)], Appendix A) that $T'_{\nu'\nu}$ is also independent of p_κ, although it is not diagonal. Inserting Equation (6.34) into the generalized Sturmian eigenproblem we obtain the generalized Sturmian secular equations:

$$\sum_\nu \left[\delta_{\nu'\nu} Z\mathcal{R}_\nu + T'_{\nu'\nu} - p_\kappa \delta_{\nu'\nu}\right] B_{\nu\kappa} = 0 \tag{6.35}$$

We note that the only thing that requires any effort to calculate in Equation (6.35) is the interelectron repulsion matrix $T'_{\nu'\nu}$ — the rest is trivial.

We shall call $T'_{\nu'\nu}$ the "energy-independent interelectron repulsion matrix". Its elements are pure numbers that depend only on the number of electrons N. Having generated $T'_{\nu'\nu}$, we can use it to calculate the properties of a large number of states for an entire isoelectronic series.

6.2 Advantages; Some illustrative examples

We have just seen the remarkable ways in which the generalized Sturmian secular equations differ from the usual secular equations that result from diagonalizing the matrix representation of the Hamiltonian of a system: We should especially notice that the eigenvalues are not energies, but values of a parameter p_κ, which is related to the energies by $E_\kappa = -p_\kappa^2/2$. In the case of Goscinskians, the configurations become pure functions $\Phi_\nu(p_\kappa \mathbf{x})$ of $p_\kappa \mathbf{x}$, i.e. p_κ acts as a scaling parameter of the space. Thus in the solution of the secular equations, an automatic scaling of the basis functions occurs: For tightly-bound states, the atomic orbitals correspond to large values of the effective charge, $Q_\nu = p_\kappa/\mathcal{R}_\nu$, and are contracted in space, whereas for loosely-bound states the orbitals are spatially diffuse. It turns out, in fact,

that the Slater exponents that are automatically obtained by solution of the generalized Sturmian secular equations are very nearly optimal. Thus, when the generalized Sturmian method is applied to atoms and atomic ions, not only is there no initial Hartree-Fock calculation, there is also no preliminary worry about what Slater exponents would be appropriate. The generalized Sturmian method using Goscinskians thus offers us a rapid and convenient method for calculating the spectra and other properties of few-electron atoms and atomic ions.

The accuracy of the method can be judged from Tables 6.2 and 6.3, comparing computed energies to experiment [Ralchenko *et al.* (2008)] and to exact nonrelativistic values calculated by Nakatsuji and coworkers [Nakashima *et al.* (2008)]. For a fixed number of electrons, the calculated values approach the exact solution to the nonrelativistic Schrödinger equation as the nuclear charge Z increases. However, in the tables, we do not see the calculation approach the experiment for the heavier elements. This is due to the fact that relativistic effects, which deepen the binding energy, rapidly become important with increasing Z. Had Nakatsuji calculated exact nonrelativistic energies for for the heavier ions, our values would have approached his, both sets of values being much less tightly bound than the experimental values.

Figures 6.1 and 6.2 illustrate this trend. However, we can notice in Tables 6.2 and 6.3, that the exact nonrelativistic value for He is more tightly bound than the experimental value. This is due to the the small role of relativistic effects and the large correction for the moving nucleus in the light helium atom. Nakatsuji and coworkers also made a calculation taking into account the motion of the nucleus, with correspondingly less tightly bound results. For 1s2s ^{1}S, adding motion of the nucleus shifts the nonrelativistic energy from -2.1460 to -2.1457, i.e. the relativistic effects deepen the binding energy by 10^{-4} Hartrees to -2.1458. In Tables 6.2 and 6.3, both Nakatsuji's values and our values neglect nuclear motion.

The Goscinskian basis is unsuitable for calculating ground states of the heliumlike isoelectronic series to high precision (see Section 6.3). These states can be calculated much more accurately using a different generalized Sturmian basis. Values calculated using 102 isoenergetic configurations based on Coulomb Sturmians are given by us in Reference [Avery and Avery (2007)], Table F.1. For the helium ground state we obtain -2.90250 Hartrees.

Table 6.2: ^1S excited state energies (in Hartrees) for the 2-electron isoelectronic series. The basis set used consisted of 40 generalized Sturmians of the Goscinski type, and the whole table was computed in a few milliseconds. Experimental values are taken from the NIST tables [Ralchenko *et al.* (2008)] (http://physics.nist.gov/asd), and the exact nonrelativistic results of Nakatsuji and coworkers [Nakashima *et al.* (2008)] are also given for comparison.

	He	Li^+	Be^{2+}	B^{3+}	C^{4+}	N^{5+}
1s2s ^{1}S	-2.1429	-5.0329	-9.1730	-14.564	-21.206	-29.098
Nakatsuji	-2.1460					
expt.	-2.1458	-5.0410	-9.1860	-14.582	-21.230	-29.131
1s3s ^{1}S	-2.0603	-4.7297	-8.5099	-13.402	-19.406	-26.521
Nakatsuji	-2.0613					
expt.	-2.0611	-4.7339	-8.5183	-13.415	-19.425	-26.548
1s4s ^{1}S	-2.0332	-4.6276	-8.2837	-13.003	-18.785	-25.629
Nakatsuji	-2.0336					
expt.	-2.0334	-4.6299	-8.2891		-18.801	-25.654
1s5s ^{1}S	-2.0210	-4.5811	-8.1806	-12.820	-18.500	-25.220
Nakatsuji	-2.0212					
expt.	-2.0210	-4.5825			-18.513	-25.241
1s6s ^{1}S	-2.0144	-4.5562	-8.1250	-12.721	-18.346	-24.998
Nakatsuji	-2.0146					
expt.	-2.0144	-4.5571				
1s7s ^{1}S	-2.0105	-4.5412	-8.0917	-12.662	-18.253	-24.865
Nakatsuji	-2.0106					
expt.	-2.0104	-4.5418				
1s8s ^{1}S	-2.0080	-4.5315	-8.0701	-12.624	-18.194	-24.779
Nakatsuji	-2.0081					
expt.	-2.0079					
1s9s ^{1}S	-2.0063	-4.5248	-8.0554	-12.598	-18.153	-24.720
Nakatsuji	-2.0064					
expt.	-2.0062					
1s10s ^{1}S	-2.0051	-4.5201	-8.0449	-12.579	-18.124	-24.678
Nakatsuji	-2.0051					
expt.	-2.0050					
1s11s ^{1}S	-2.0042	-4.5166	-8.0371	-12.566	-18.102	-24.647
Nakatsuji	-2.0042					
expt.	-2.0041					
1s12s ^{1}S	-2.0034	-4.5140	-8.0312	-12.555	-18.086	-24.624
Nakatsuji	-2.0036					
expt.	-2.0034					

Table 6.3: ^{3}S excited state energies calculated with 36 Goscinskians. The calculation of similar tables for ^{1}P, ^{3}P, ^{1}D, ^{3}D, doubly excited autoionizing states, etc., is equally easy, rapid, and of comparable accuracy. Tables are given in Chapters 3 and 4 in [Avery and Avery (2007)], but may easily be reproduced using our programs, as shown in Tutorial 1 on [Avery and Avery (2006)].

	He	Li^+	Be^{2+}	B^{3+}	C^{4+}	N^{5+}
1s2s ^{3}S	-2.1736	-5.1079	-9.2937	-14.730	-21.417	-29.353
Nakatsuji	-2.1752					
expt.	-2.1750	-5.1109	-9.2983	-14.738	-21.429	-29.375
1s3s ^{3}S	-2.0682	-4.7504	-8.5442	-13.450	-19.466	-26.594
Nakatsuji	-2.0687					
expt.	-2.0685	-4.7522	-8.5480	-13.457	-19.478	-26.614
1s4s ^{3}S	-2.0363	-4.6360	-8.2983	-13.023	-18.811	-25.661
Nakatsuji	-2.0365					
expt.	-2.0363	-4.6373	-8.3015	-13.030	-18.822	-25.680
1s5s ^{3}S	-2.0225	-4.5854	-8.1880	-12.831	-18.514	-25.237
Nakatsuji	-2.0226					
expt.	-2.0224	-4.5862	-8.1905		-18.524	-25.254
1s6s ^{3}S	-2.0153	-4.5586	-8.1293	-12.728	-18.354	-25.008
Nakatsuji	-2.0154					
expt.	-2.0152	-4.5592			-18.364	
1s7s ^{3}S	-2.0111	-4.5427	-8.0944	-12.667	-18.259	-24.872
Nakatsuji	-2.0111					
expt.	-2.0109	-4.5431			-18.268	
1s8s ^{3}S	-2.0084	-4.5325	-8.0719	-12.627	-18.197	-24.784
Nakatsuji	-2.0084					
expt.	-2.0082	-4.5328			-18.206	
1s9s ^{3}S	-2.0066	-4.5255	-8.0567	-12.600	-18.156	-24.724
Nakatsuji	-2.0066					
expt.	-2.0064					
1s10s ^{3}S	-2.0053	-4.5206	-8.0458	-12.581	-18.126	-24.681
Nakatsuji	-2.0053					
expt.	-2.0051					
1s11s ^{3}S	-2.0044	-4.5170	-8.0378	-12.567	-18.103	-24.649
Nakatsuji	-2.0044					
expt.						
1s12s ^{3}S	-2.0035	-4.5142	-8.0317	-12.556	-18.087	-24.625
Nakatsuji	-2.0037					
expt.						

6.2.1 *The large-Z approximation: Restriction of the basis set to an $\mathcal{R}$-block*

One of the great advantages of the generalized Sturmian method when it is applied to atoms and atomic ions is that it leads naturally to an approximation that allows us to write down the energies of atomic states with so little effort that the calculation can literally be carried out on the back of an envelope! We call this approximation the *Large-Z Approximation.*

If interelectron repulsion is entirely neglected, i.e. when disregarding the second term in Equation (6.19), the calculated energies E_κ of Ψ_κ become those of a set of N completely independent electrons moving in the field of the bare nucleus:

$$E_\kappa = -\frac{p_\kappa^2}{2} \longrightarrow -\frac{1}{2}Z^2{\mathcal{R}_\nu}^2 = -\frac{Z^2}{2n_1^2} - \frac{Z^2}{2n_2^2} - \cdots - \frac{Z^2}{2n_N^2} \tag{6.36}$$

Equation (6.36) is *not* the large-Z approximation:

> In the large-Z approximation, we do not neglect interelectron repulsion, but we restrict the basis set to those Goscinskian configurations that would be degenerate if interelectron repulsion were entirely neglected, i.e., we restrict the basis to a set of configurations all of which correspond to the same value of $\mathcal{R}_\nu$.

In that case, the term $\delta_{\nu'\nu}Z\mathcal{R}_\nu$ in (6.19) is a multiple of the identity matrix, and the eigenvectors $B_{\nu\kappa}$ are the same as those that would be obtained by diagonalizing the energy-independent interelectron repulsion matrix $T'_{\nu'\nu}$, since eigenfunctions are unchanged by adding a multiple of the unit matrix.

$$\sum_\nu \left[T'_{\nu'\nu} - \lambda_\kappa \delta_{\nu'\nu}\right] B_{\nu\kappa} = 0 \tag{6.37}$$

The roots are shifted by an amount equal to the constant by which the identity matrix is multiplied:

$$p_\kappa = Z\mathcal{R}_\nu + \lambda_\kappa = Z\mathcal{R}_\nu - |\lambda_\kappa| \tag{6.38}$$

and the energies become

Table 6.4: Roots of the ground state $\mathcal{R}$-block of the interelectron repulsion matrix for the He-like, Li-like, Be-like, B-like and C-like isoelectronic series.

Li-like $\lvert\lambda_\kappa\rvert$	term	Be-like $\lvert\lambda_\kappa\rvert$	term	B-like $\lvert\lambda_\kappa\rvert$	term	C-like $\lvert\lambda_\kappa\rvert$	term
0.681870	^{2}S	0.986172	^{1}S	1.40355	^{2}P	1.88151	^{3}P
0.729017	^{2}P	1.02720	^{3}P	1.44095	^{4}P	1.89369	^{1}D
		1.06426	^{1}P	1.47134	^{2}D	1.90681	^{1}S
		1.09169	^{3}P	1.49042	^{2}S	1.91623	^{5}S
		1.10503	^{1}D	1.49395	^{2}P	1.995141	^{3}D
		1.13246	^{1}S	1.52129	^{4}S	1.96359	^{3}P
				1.54037	^{2}D	1.98389	^{3}S
				1.55726	^{2}P	1.98524	^{1}D
						1.99742	^{1}P
						2.04342	^{3}P
He-like						2.05560	^{1}D
0.441942	^{1}S					2.07900	^{1}S

Table 6.5: Roots of the ground state $\mathcal{R}$-block of the interelectron repulsion matrix $T'_{\nu'\nu}$ for the N-like, O-like, F-like and Ne-like isoelectronic series.

N-like $\lvert\lambda_\kappa\rvert$	term	O-like $\lvert\lambda_\kappa\rvert$	term	F-like $\lvert\lambda_\kappa\rvert$	term	Ne-like $\lvert\lambda_\kappa\rvert$	term
2.41491	^{4}S	3.02641	^{3}P	3.68415	^{2}P	4.38541	^{1}S
2.43246	^{2}D	3.03769	^{1}D	3.78926	^{2}S		
2.44111	^{2}P	3.05065	^{1}S				
2.49314	^{4}P	3.11850	^{3}P				
2.52109	^{2}D	3.14982	^{1}P				
2.53864	^{2}S	3.24065	^{1}S				
2.54189	^{2}P						
2.61775	^{2}P						

$$E_\kappa = -\frac{1}{2}p_\kappa^2 = -\frac{1}{2}(Z\mathcal{R}_\nu - |\lambda_\kappa|)^2 \tag{6.39}$$

Since the roots λ_κ are always negative, we may use the form $-|\lambda_\kappa|$ in place of λ_κ to make explicit the fact that interelectron repulsion reduces the binding energies, as of course it must. The roots λ_κ are pure numbers that can be calculated once and for all and stored. Values of these roots for N=2,3,...,10 are shown in Tables 6.4 and 6.5. The spectroscopic terms to which the roots correspond are also shown. From the roots, a great deal of information about atomic states can be found with almost no effort: Given the values of the principle quantum numbers $n_1, n_2, ..., n_N$, from which $\mathcal{R}_\nu$ can easily be found, and given the value of $|\lambda_\kappa|$, which can be looked up in a table, the calculation of the energies for the entire isoelectronic series is completely effortless!

The eigenfunctions corresponding to the spectroscopic terms in Tables 6.4 and 6.5 are symmetry adapted Russel-Saunders states and can be used as basis functions for more exact calculations. The classification is done automatically by the method discussed in [Avery and Avery (2007)], Sections 3.4 and 3.5. Tutorial 2 on our website `sturmian.kvante.org` [Avery and Avery (2006)] shows in detail how to do this.

6.2.2 *Validity of the large-Z approximation*

In Figure 6.1, the large-Z approximation energy $E_\kappa = -\frac{1}{2}(Z\mathcal{R}_\nu - |\lambda_\kappa|)^2$ for the lowest triplet states of the helium-like isoelectronic series is plotted against spectroscopically determined energies. In order to better see the details, we plot E_κ/Z^2. Figure 6.2 shows E_κ/Z^2 for the ground state of the six-electron isoelectronic series.

As the nuclear charge Z increases, the energies and wave functions calculated with the large-Z approximation approach the exact solutions to the non-relativistic Schrödinger equation. However, relativistic effects begin to be pronounced at around $Z = 10$, and become progressively more so as Z increases. Therefore the calculated values first approach the experimental ones, but begin to differ as relativity becomes more and more important.

It is possible to make a rough correction for the relativistic effect on the energies by multiplying them by an easily-calculated factor $f_\kappa(Z)$, so that E_κ becomes

$$-\frac{1}{2}f_\kappa(Z)(Z\mathcal{R}_\nu + \lambda_\kappa)^2 \tag{6.40}$$

The correction factor $f_\kappa(Z)$ is the ratio between the relativistic and nonrelativistic energies of a configuration, assuming interelectron repulsion to be completely neglected such that the energy is equal to that of N independent electrons moving in the field of the nucleus.

In the nonrelativistic case, the energy of a hydrogenlike spin-orbital is given by $-\frac{Z^2}{2n^2}$, and thus the total energy of an N-electron configuration is $-\frac{1}{2}Z^2{\mathcal{R}_\nu}^2$.

In the relativistic case, the exact solution to the Dirac equation for hydrogenlike atoms can be found in [Akhiezer and Berestetskii (1965)], or in [Avery and Avery (2007)], Equations (7.35) through (7.40). The ratio of the relativistic energy $E_{\rm rel}$ and the nonrelativistic energy $E_{\rm nonrel}$ for a multiconfigurational state

$$\Psi_\kappa = \sum_\nu \Phi_\nu B_{\nu\kappa} \tag{6.41}$$

is

$$f_\kappa(Z) = \frac{E_{\rm rel}}{E_{\rm nonrel}} = \frac{\sum_\nu B_{\nu\kappa}^2 \langle\Phi_\nu|\mathsf{H}_0|\Phi_\nu\rangle_{\rm rel}}{\sum_\nu B_{\nu\kappa}^2 \langle\Phi_\nu|\mathsf{H}_0|\Phi_\nu\rangle_{\rm nonrel}} = \frac{\sum_\nu B_{\nu\kappa}^2 \langle\Phi_\nu|\mathsf{H}_0|\Phi_\nu\rangle_{\rm rel}}{-\frac{1}{2}Z^2\sum_\nu B_{\nu\kappa}^2{\mathcal{R}_\nu}^2} \tag{6.42}$$

Here, H_0 is a sum of one-electron Hamiltonian operators corresponding to single electrons moving in the field of the bare nucleus, i.e. interelectron repulsion is completely neglected.

In the figures, the lines are calculated in the large-Z approximation. The upper (dashed) line is not corrected for relativistic effects, while the lower (solid) line is corrected. The dots are experimental values of the energies taken from the NIST Atomic Spectra Database [Ralchenko *et al.* (2008)]. It can be seen from Figures 6.1 and 6.2 that agreement between the energies calculated from the large-Z approximation and experimental energies become progressively better as Z increases, provided that the rough relativistic correction is made.

We note that the large-Z approximation, despite its great simplicity, well approximates non-relativistic energies: Even for modest values of nuclear

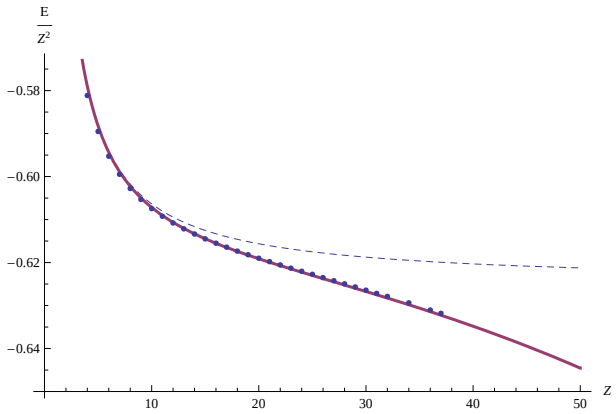


Fig. 6.1 This figure shows energies for the lowest ^{3}S state of the helium-like isoelectronic series, divided by Z^2 to make the details easier to see for large Z. The values are calculated in the large-Z approximation, which here limits the basis to a single configuration. The lower (solid) line is corrected for relativistic effects as discussed in the text; the dots indicate experimental values from the NIST tables. It is easy to visually verify that, for $Z > 10$, the relativistic correction is much larger than calculational errors due to the large-Z approximation.

charge, the error of the large-Z approximation is much smaller than the error due to neglecting relativity. Further, relativistic effects may be accounted for by means of an easily calculated factor, yielding energies that correspond well with experiment.

The second example presented here (the ground state of the carbonlike isoelectronic series) is a case not easily approximated using a small number of Goscinskian basis functions, because interelectron repulsion plays a large role. Nevertheless, it can be seen that even in this somewhat difficult case, the large-Z approximation gives surprisingly reasonable results. The large-Z approximation is not only extremely simple, but it is characterized by a small number of parameters – the roots of the interelectron repulsion matrix. These roots are dimensionless and independent of energy and of nuclear charge. They can be calculated once and for all, and they contain information concerning many states of the entire isoelectronic series. Once the roots are obtained, calculating approximate atomic energies, and a number of other properties, become tasks that can be carried out by pen and paper.

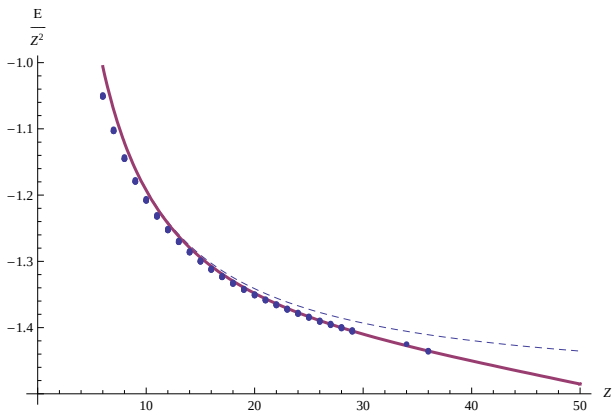


Fig. 6.2 The ground state of the carbon-like isoelectronic series. As Z grows, the approximation approaches the exact solution to the nonrelativistic Schrödinger equation. Due to the increased role of interelectron repulsion in the carbon-like series, this takes longer than for the helium-like series. However, at around $Z = 18$, the inaccuracy of the large-Z approximation becomes smaller than the relativistic correction.

6.2.3 *Core ionization energies*

The large-Z approximation can be used to estimate a number of additional properties. For example, using the approximation, we can calculate by hand the core-ionization energies, i.e. the energies required to remove an electron from the inner shell of an atom. From (6.39) we can see that this energy will be given by

$$\Delta E = \frac{1}{2}\left[(Z\mathcal{R}_\nu - |\lambda_\kappa|)^2 - (Z\mathcal{R}_\nu{}' - |\lambda'_\kappa|)^2\right] \tag{6.43}$$

where the unprimed quantities refer to the original ground state, while the primed quantities refer to the core-ionized states. Since

$$\mathcal{R}_\nu{}^2 - \mathcal{R}_\nu{}'^2 = 1 \tag{6.44}$$

Equation (6.43) can be written in the form

$$\Delta E - \frac{Z^2}{2} = Z\left[\mathcal{R}_\nu{}'|\lambda'_\kappa| - \mathcal{R}_\nu|\lambda_\kappa|\right] + \frac{|\lambda_\kappa|^2 - |\lambda'_\kappa|^2}{2} \tag{6.45}$$

Thus we can see that within the framework of the large-Z approximation, the quantity $\Delta E - Z^2/2$ is linear in Z for an isoelectronic series. This quantity represents the contribution of interelectron repulsion to the core ionization energy, since if interelectron repulsion is completely neglected, the core ionization energy is given by $\Delta E = Z^2/2$. Core ionization energies

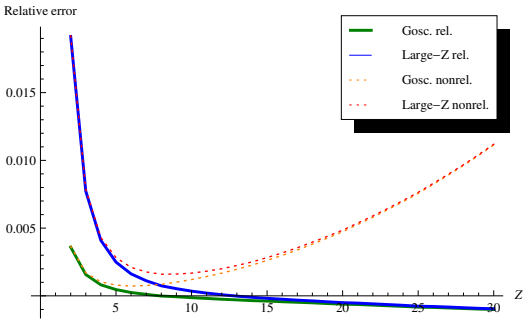


Fig. 6.3 Ground state relative errors $\frac{E_{calc}-E_{exp}}{E_{exp}}$ compared to experiment for the helium-like isoelectronic series. The large-Z approximation energies $-\frac{1}{2}(Z\sqrt{2}-.441942)^2$ are compared to results using a fuller Goscinskian basis. The two dotted lines are the nonrelativistic values, while the solid lines are corrected for relativistic effects using Equation (6.42). For very large values of Z, errors due to quantum electrodynamic effects cause a systematic overestimation of binding energy.

.

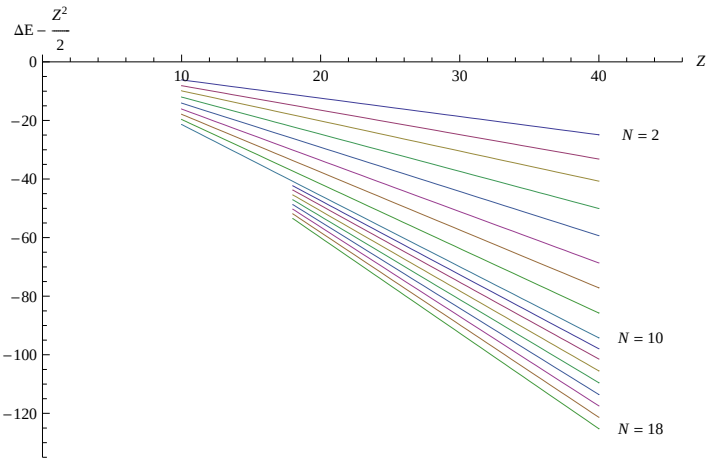


Fig. 6.4 For isoelectronic series, Equation (6.45) indicates that within the large-Z approximation, the quantity $\Delta E - Z^2/2$ is exactly linear in Z, as is illustrated above.

calculated from Equations (6.43)-(6.45) are shown in Figure 6.4. Between $N = 10$ and $N = 18$ the lines in the figure become more closely spaced. This is because a new shell starts to fill at $N = 11$. A table showing qualitative agreement between experiment and the core-ionization energies

calculated in the large-Z approximation is given in Chapter 5 of [Avery and Avery (2007)]. Detailed calculations can be found on our website [Avery and Avery (2006)] in Tutorials 3 and 5.

6.3 Limitations of the method; Prospects for the future

We mentioned above that the generalized Sturmian method using Goscinskian configurations offers a very rapid and convenient method for calculating the spectra, wave functions and other properties of few-electron atoms and atomic ions. But why is the method limited to systems with a small number of electrons? The reason for this is that Goscinskian configurations are solutions to Equation (6.25) with V_0 chosen as the Coulomb attraction of the bare nucleus, as shown in Equation (6.26). As the number of electrons N becomes large, this zeroth-order potential becomes progressively more unrealistic, because the effects of interelectron repulsion become progressively more important.

How can we correct this defect? One way to extend the range of the method is to use a V_0 in Equation (6.25) that in some form includes interelectron repulsion effects. This will make it less straightforward to obtain the generalized Sturmian configurations Φ_ν, depending on the complexity of the chosen V_0, in general requiring a self-consistent field iteration. However, the useful properties of the generalized Sturmian basis are retained, and the extra initial work would lead to improved convergence.

Another possibility is to extend the method by using a basis set consisting of isoenergetic configurations

$$\Phi_\nu(\mathbf{x}) = |\varphi_{\zeta_1}\varphi_{\zeta_2}...\varphi_{\zeta_N}| \tag{6.46}$$

constructed from orbitals satisfying

$$\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(r_j)\right]\varphi_\zeta(\mathbf{x}_j) = 0 \tag{6.47}$$

where $v(r_j)$ is the nuclear attraction potential, corrected by a repulsive potential due to the core electrons:

$$v(r_j) = -\frac{Z}{r_j} + v_c(r_j) \tag{6.48}$$

This introduces interelectron repulsion effects even earlier in the calculations. The potential $v_c(r_j)$ can be found by performing a fast preliminary

calculation using Goscinskian configurations. From this preliminary calculation, a spherically-averaged core density, $\rho(r_j)$ can be obtained, and from this $v_c(r_j)$ may be calculated by means of the relationship

$$v_c(r_j) = \int_0^\infty dr'_j \; r'^2_j \; \rho(r'_j) \frac{1}{r_>}, \qquad r_> \equiv Max[r_j, r'_j] \tag{6.49}$$

The orbitals $\varphi_\zeta(\mathbf{x}_j)$ can be built up from Coulomb Sturmians, so that (6.47) becomes:

$$\sum_\mu \left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(r_j) \right] \chi_\mu(\mathbf{x}_j) C_{\mu,\zeta} = 0 \tag{6.50}$$

Multiplying from the left by a conjugate Coulomb Sturmian, we obtain:

$$\begin{aligned} 0 &= \sum_\mu \int d^3x_j \chi^*_{\mu'}(\mathbf{x}_j) \left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(r_j) \right] \chi_\mu(\mathbf{x}_j) C_{\mu,\zeta} \\ &= \sum_\mu \left[k^2 \delta_{\mu'\mu} + \beta_\nu \int d^3x_j \chi^*_{\mu'}(\mathbf{x}_j) v(r_j) \chi_\mu(\mathbf{x}_j) \right] C_{\mu,\zeta} \\ &= \sum_\mu \left[k^2 \delta_{\mu'\mu} - k\beta_\nu t_{\mu',\mu} \right] C_{\mu,\zeta} \end{aligned} \tag{6.51}$$

or

$$\sum_\mu \left[t_{\mu',\mu} - \frac{k}{\beta_\nu} \delta_{\mu'\mu} \right] C_{\mu,\zeta} = 0 \tag{6.52}$$

where

$$t_{\mu',\mu} \equiv -\frac{1}{k} \int d^3x_j \chi^*_{\mu'}(\mathbf{x}_j) v(r_j) \chi_\mu(\mathbf{x}_j) \tag{6.53}$$

After solving Equation (6.53) to obtain the coefficients $C_{\mu,\zeta}$, we can next use the isoenergetic configurations $\Phi_\nu(\mathbf{x}) = |\varphi_{\zeta_1}\varphi_{\zeta_2}...\varphi_{\zeta_N}|$ as basis functions for solving the Schrödinger equation for an atom or atomic ion. This can be written in the form

$$\left[\sum_{j=1}^N \left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} \right) + \mathsf{V}(\mathbf{x}) \right] \Psi_\kappa(\mathbf{x}) = 0 \tag{6.54}$$

with

$$\mathsf{V}(\mathbf{x}) = -\sum_{j=1}^N \frac{Z}{r_j} + \sum_{i>j}^N \sum_{j=1}^N \frac{1}{r_{ij}} \tag{6.55}$$

and with

$$E_\kappa = -\sum_{j=1}^{N} \frac{k^2}{2} = -\frac{Nk^2}{2} \tag{6.56}$$

Thus we write

$$\Psi_\kappa(\mathbf{x}) = \sum_\nu \Phi_\nu(\mathbf{x}) B_{\nu,\kappa} \tag{6.57}$$

Substituting this into the N-electron Schrödinger equation, and taking the scalar product with a conjugate configuration, we obtain the secular equations:

$$\sum_\nu \int dx\ \Phi^*_{\nu'}(\mathbf{x}) \left[\sum_{j=1}^{N} \left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} \right) + \mathsf{V}(\mathbf{x}) \right] \Phi_\nu(\mathbf{x}) B_{\nu,\kappa} = 0 \tag{6.58}$$

We next introduce the k-independent matrix

$$\mathcal{S}_{\nu',\nu} \equiv \frac{1}{k^2} \int dx\ \Phi^*_{\nu'}(\mathbf{x}) \sum_{j=1}^{N} \left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} \right) \Phi_\nu(\mathbf{x}) \tag{6.59}$$

which can be interpreted as a generalized Shibuya-Wulfman matrix, and another k-independent matrix:

$$T_{\nu',\nu} \equiv -\frac{1}{k} \int dx\ \Phi^*_{\nu'}(\mathbf{x}) \mathsf{V}(\mathbf{x}) \Phi_\nu(\mathbf{x}) \tag{6.60}$$

In terms of these matrices, the secular equations become:

$$\sum_\nu \left[T_{\nu',\nu} - k\mathcal{S}_{\nu',\nu} \right] B_{\nu,\kappa} = 0 \tag{6.61}$$

This gives us a spectrum of k-values from which the energies of the various states, $E_\kappa = -Nk^2/2$, can be obtained. It seems quite likely that this procedure would allow the generalized Sturmian method for atoms and atomic ions to be extended to larger values of N. Some steps in this direction have already been taken by us and by Professor Gustavo Gasaneo and his students at Universidad Nacional del Sur in Argentina.

What developments are necessary in order to apply the generalized Sturmian method to complex chemical problems? Once we have found a generalized Sturmian basis that converges well, most of the standard techniques in quantum chemistry can be employed in the same way that they are

currently used with bases obtained from initial Hartree-Fock calculations. Two obvious steps are to use the frozen core approximation to factor out correlation of core electrons, and use standard perturbation theory based techniques to reduce the computational efforts necessary for configuration interaction. Using the generalized Sturmian method with for example Coupled Cluster methods requires some work, but may be well worth the effort due to improved convergence properties compared to using Hartree-Fock based configurations.

6.3.1 *Can the generalized Sturmian method be applied to N-electron molecules?*

If we wish to use a many-center Coulomb Sturmian basis set to treat molecules, we can introduce the notation

$$\chi_\tau(\mathbf{x}) \equiv \chi_\mu(\mathbf{x} - \mathbf{X}_a) \tag{6.62}$$

where

$$\tau \equiv (a, l, m, n) \tag{6.63}$$

In a molecule, the one-electron nuclear attraction potential has the form:

$$v(\mathbf{x}_j) = -\sum_a \frac{Z_a}{|\mathbf{x}_j - \mathbf{X}_a|} \tag{6.64}$$

We can build up solutions to the one-electron equation

$$\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + v(\mathbf{x}_j)\right] \varphi_\zeta(\mathbf{x}_j) = 0 \tag{6.65}$$

from superpositions of many-center Coulomb Sturmians:

$$\varphi_\zeta(\mathbf{x}_j) = \sum_\tau \chi_\tau(\mathbf{x}_j) C_{\tau,\zeta} \tag{6.66}$$

Thus we have

$$\sum_\tau \left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + v(\mathbf{x}_j)\right] \chi_\tau(\mathbf{x}_j) C_{\tau,\zeta} = 0 \tag{6.67}$$

Taking the scalar product with a conjugate Coulomb Sturmian yields:

$$\sum_\tau \int d^3x_j \; \chi^*_{\tau'}(\mathbf{x}_j) \left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + v(\mathbf{x}_j)\right] \chi_\tau(\mathbf{x}_j) C_{\tau,\zeta} = 0 \tag{6.68}$$

If we let

$$\mathfrak{S}_{\tau'\tau} \equiv \frac{1}{k^2} \int d^3x_j \; \chi^*_{\tau'}(\mathbf{x}_j) \left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} \right] \chi_\tau(\mathbf{x}_j) \tag{6.69}$$

and

$$\mathfrak{W}_{\tau'\tau} \equiv -\frac{1}{k} \int d^3x_j \; \chi^*_{\tau'}(\mathbf{x}_j) v(\mathbf{x}_j) \chi_\tau(\mathbf{x}_j) \tag{6.70}$$

then the one-electron secular equations take the form:

$$\sum_\tau \left[\mathfrak{W}_{\tau'\tau} - \frac{k}{\beta_\nu} \mathfrak{S}_{\tau'\tau} \right] C_{\tau\zeta} = 0 \tag{6.71}$$

The integrals $\mathfrak{S}_{\tau'\tau}$ are the well-studied Shibuya-Wulfman integrals, which can be generated using a variety of algorithms.

Historical note: The Shibuya-Wulfman integrals were first introduced by T. Shibuya and C. Wulfman in 1965 in connection with their famous momentum-space treatment of many-center one-electron problems[T. and C.E. (1965)]. These integrals can conveniently be evaluated in momentum space using the Fock projection, which establishes a relationship between hyperspherical harmonics and the Fourier transforms of Coulomb Sturmians. The problem of evaluating these integrals, as well as many other integrals needed in molecular problems, can then be converted into a problem of hyperangular integration [Avery (1989, 1998); Caligiana (2003); Avery and Avery (2007)].

It can be shown that $\mathfrak{W}_{\tau'\tau}$ and $\mathfrak{S}_{\tau'\tau}$ are related through the sum rule [Koga and Matsuhashi (1987, 1988)]:

$$\mathfrak{W}_{\tau'\tau''} = \sum_\tau \mathfrak{S}_{\tau'\tau} \frac{Z_a}{n} \mathfrak{S}_{\tau\tau''} \tag{6.72}$$

In Equation (6.72), $\tau = (a, \mu)$ runs over all the atoms a in the molecule and the full Coulomb Sturmian basis set centered at $\mathbf{X}_a$. If the basis set is truncated, the relationship is only approximate. Solutions to

the two-center 1-electron Schrödinger equation using the method just described have been studied by Aquilanti and his group in Perugia [Aquilanti *et al.* (1996b)],[Aquilanti *et al.* (1996a)],[Caligiana (2003)] and by Koga and his group in Japan [Koga and Matsuhashi (1987)],[Koga and Matsuhashi (1988)].

We now introduce the N-electron configurations of the form

$$\Phi_\nu(\mathbf{x}) = |\varphi_{\zeta_1}\varphi_{\zeta_2}...\varphi_{\zeta_N}| \tag{6.73}$$

where the molecular orbitals φ_{ζ_j} are solutions to (6.71). We would like to use these configurations to build up solutions to the N-electron Schrödinger equation

$$\left[\sum_{j=1}^{N}\left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right) + \mathsf{V}(\mathbf{x})\right]\Psi_\kappa(\mathbf{x}) = 0 \quad \text{with} \quad E_\kappa = -\frac{Nk^2}{2} \tag{6.74}$$

where

$$\mathsf{V}(\mathbf{x}) = -\sum_{j=1}^{N}\sum_{a}\frac{Z_a}{|\mathbf{x}_j - \mathbf{X}_a|} + \sum_{i>j}^{N}\sum_{j=1}^{N}\frac{1}{r_{ij}} \tag{6.75}$$

Then in a manner exactly similar to equations (6.54)-(6.58) we are led to secular equations of the form

$$\sum_{\nu}\left[T_{\nu',\nu} - k\mathcal{S}_{\nu',\nu}\right]B_{\nu,\kappa} = 0 \tag{6.76}$$

the only difference being that in the calculation of $T_{\nu',\nu}$, the molecular potential is used in place of the atomic potential.

In the case of diatomic molecules, we begin by picking a value of the parameter $s = kR$, where R is the interatomic distance and k is the exponent of the Coulomb Sturmian basis set. Neither R nor k is known at this point, but only their product s. For the diatomic case, all of the integrals involved in Equations (6.71) and (6.61) are pure functions of s. Having chosen s, we can thus solve the one-electron secular equations and obtain the coefficients $C_{\tau\zeta}$ and the spectrum of ratios k/β_ν. We are then able to solve Equation (6.61), which gives us the eigenvectors $B_{\nu\kappa}$ as well as a spectrum of k-values, and thus energies $-Nk^2/2$. From a k-value, we also get the unscaled distance $R = s/k$. We repeat the procedure for a range of

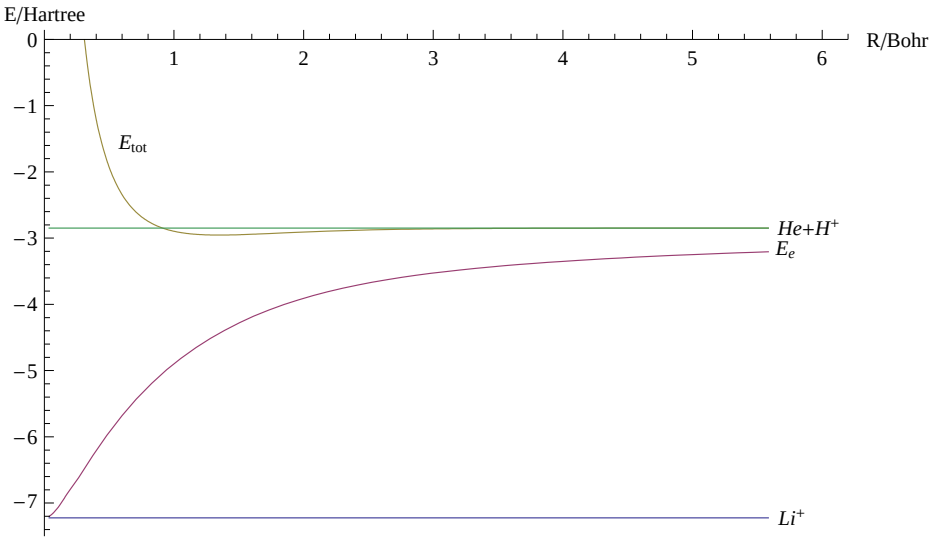


Fig. 6.5 This figure shows the electronic energy E_e and the total energy E_{tot} of the HeH^+ ion as a function of the internuclear separation $R = s/k$. The calculation was performed with a single configuration using a one-electron basis set consisting of 3 Coulomb Sturmians on each center. For $R \to 0$, the electronic energy approaches the energy calculated for the Li^+ ion using the generalized Sturmian method with a single configuration [Avery and Avery (2007)]. In the separated atom region, the total energy approaches that of He when calculated in the same way. Our calculation exhibits a shallow minimum at $R = 1.35$ Bohrs, which can be compared to the equilibrium bond length of 1.3782 Bohrs resulting from a HF/STO-3G calculation quoted by Szabo and Ostlund [Szabo and Ostlund (1996)], and with the value of 1.46 Bohrs obtained in a benchmark calculation by Wolniewicz [Wolniewicz (1965)]. Since our pilot calculation uses only one configuration, it makes sense that we obtain a result comparable to the Hartree-Fock calculation.

s-values and interpolate to find the solutions as functions of R. Figure 6.5 shows our pilot calculation on the HeH^+ two-electron molecular ion using the method described above. This is an extremely simple calculation, using only one configuration, but we are actively working to explore the method further. We chose HeH^+ for the pilot calculation rather than H_2 because, as is well known, the correct dissociation curve for H_2 needs at least two configurations.

In the case of polyatomic molecules, one can choose a set of angles between the nuclei; these are left fixed under scaling of the coordinate system. The procedure is then similar to that described for the diatomic case.

6.4 Discussion

In this chapter, we have concentrated on the use of isoenergetic configurations of the Goscinskian type. However, the generalized Sturmian method can be applied to wide variety of problems, including cases where the masses are unequal and the force field very different from Coulomb interactions. All that is needed is a set of isoenergetic solutions to Equation (6.9), where V_0 determines the shapes of the resulting generalized Sturmian basis functions.

We have performed calculations based on isoenergetic configurations formed from Coulomb Sturmians, as is discussed in our recent book ([Avery and Avery (2007)], Appendix F). Basis sets of this type have the advantage that we have rigorous proof of their completeness. Unlike most other methods, this thus gives assurance that we can approximate the exact solution to the Schrödinger equation with arbitrary precision, given arbitrary computation time. In Section 6.3, we outline how to build a generalized Sturmian basis for molecules. Professor Gustavo Gasaneo and his coworkers are exploring Sturmian bases that include interelectron repulsion effects as well as Sturmians that include the continuum.

We have seen that the generalized Sturmian method using Goscinskian configurations as a basis offers an extremely rapid and convenient way of calculating atomic spectra and other properties of few-electron atoms and ions. We obtain good solutions to many states at once, using only a very small basis and without any need for SCF calculations. The wave function is in a form that is easy to interpret by inspection or to analyze automatically by computer (see Tutorial 2 at [Avery and Avery (2006)]). Furthermore, $T'_{\nu'\nu}$, the matrix which represents interelectron repulsion in a basis of Goscinskian configurations, consists of pure numbers of universal applicability. The matrix depends only on the number of electrons in an atom or ion, and it is independent of energy and nuclear charge. The *energy-independent interelectron repulsion matrix*, $T'_{\nu'\nu}$, can be calculated once and for all, and used in a wide variety of cases.

The results shown in Section 2 illustrate the degree of speed and accuracy that can be expected for few-electron atoms and ions. Neutral atoms are the worst case when using Goscinskians. However, in the neutral helium atom, where Nakatsuji's results [Nakashima *et al.* (2008)] are available, our results agree well with his as can be seen in Tables 6.2 and 6.3. Had Nakatsuji and

coworkers made calculations on the whole isoelectronic series, agreement would be progressively better for the heavier ions in the series.

We find that in order to obtain good agreement with experiment, it is necessary to include relativistic effects. For the few-electron systems treated here, the crude relativistic correction of Equation (6.42) gives very good results. For the two-electron isoelectronic series, the ground state was obtained with relative error compared to experiment of $3.5 \cdot 10^{-3}$ for $Z = 2$ (the worst case) to roughly 10^{-6} for $Z \geq 8$, and excited states were obtained with relative errors between 10^{-4} and 10^{-6}. The complete calculation of all the states ([Avery (2008)], Chapter 4) required only 77ms of computation. It should be noted, that for very large values of Z, quantum electrodynamic effects become important, and neglecting them will cause an overestimation of the binding energies. If more precision is required, we can treat the system by means of the Dirac-Coulomb equation. Calculations using a fully relativistic analogue to the Goscinskian configurations can be found in Chapter 7 of our book [Avery and Avery (2007)].

As the number of electrons grow, there is a decrease in accuracy when using similar computation time. Already for five-electron systems, the calculations in Chapter 4 of [Avery (2008)] yield ground states for the $Z = N$ case with less accuracy than the Hartree-Fock limit. This suggests to us that in order to solve systems with many electrons accurately, while retaining efficiency, we need Sturmian basis functions that incorporate interelectron repulsion in $\mathsf{V}_0(\mathbf{x})$, as is discussed in Section 6.3.

The generalized Sturmian method using Goscinskians leads to an extremely simple and convenient approximation, the *large-Z approximation*, which is described in Section 6.2. The approximation leads to a remarkably simple closed-form expression, $E_\kappa = -\frac{1}{2}(Z\mathcal{R}_\nu - |\lambda_\kappa|)^2$, for the energies of states in terms of the appropriate roots of the energy-independent interelectron repulsion matrix.

As the name suggests, the large-Z approximation is not very accurate when $Z = N$, especially in the case of ground states. It underestimates the binding energy of the ground state of neutral helium by 2% and of neutral argon by 5% (Figure 4.5 in [Avery (2008)]), but it improves rapidly with increasing $Z - N$ (Figure 4.3 in [Avery (2008)]). For excited states of few-electron atoms, the large-Z approximation gives surprisingly good results even for modest values of $Z - N$. Given the interelectron repulsion roots

λ_κ, which are dimensionless quantities that depend only on the number of electrons and can be precalculated, we can calculate electronic states for entire isoelectronic series with a pencil and a scrap of paper.

It is our hope that in the future the method may be extended to give accurate calculations for atoms where interelectron repulsion effects are comparable in importance to nuclear attraction. We are also in the process of extending the method to molecules.

Bibliography

Akhiezer, A. and Berestetskii, V. (1965). *Quantum Electrodynamics* (Interscience, New York).

Aquilanti, V., Cavalli, S., Coletti, C. and Grossi, G. (1996a). Alternative Sturmian bases and momentum space orbitals; an application to the hydrogen molecular ion, *Chem. Phys* **209**, p. 405.

Aquilanti, V., Cavalli, S., D., D. F. and Grossi, G. (1996b). Hyperangular momentum: Applications to atomic and molecular science, in C. Tsipis, V. Popov, D. Herschbach and J. Avery (eds.), *New Methods in Quantum Theory* (Kluwer Academic Publishers, Dordrecht).

Avery, J. (1989). *Hyperspherical Harmonics; Applications in Quantum Theory* (Kluwer Academic Publishers, Dordrecht).

Avery, J. (1998). A formula for angular and hyperangular integration, *J. Math. Chem.* **24**, p. 169.

Avery, J. (2003). Many-center coulomb Sturmians and Shibuya-Wulfman integrals, *Int. J. Quantum Chem.* .

Avery, J. (2008). *The Generalized Sturmian Method: Development, Implementation and Applications in Atomic Physics*, Master's thesis, Department of Computer Science, University of Copenhagen (DIKU), available at `http://sturmian.kvante.org/thesis/speciale.pdf` and `http://sturmian.kvante.org/thesis/sturmian-1.0.pdf`.

Avery, J. and Avery, J. (2005). Autoionizing states of atoms calculated using generalized Sturmians, *Adv. Quant. Chem.* **49**, pp. 103–118.

Avery, J. and Avery, J. (2006). The Generalized Sturmian Library, Available at `http://sturmian.kvante.org`.

Avery, J. and Avery, J. (2007). *Generalized Sturmians and Atomic Spectra* (World Scientific, New York).

Avery, J. and Herschbach, D. R. (1992). Hyperspherical Sturmian basis functions, *Int. J. Quantum Chem.* .

Caligiana, A. (2003). *Sturmian Orbitals in Quantum Chemistry*, Ph.D. thesis, University of Perugia, Italy.

Goscinski, O. (2002). Conjugate eigenvalue problems and generalized Sturmians, *Advances in Quantum Chemistry* **41**, pp. 51–56, doi:10.1016/

S0065-3276(02)41046-5.

Koga, T. and Matsuhashi, T. (1987). Sum rules for nuclear attraction integrals over hydrogenic orbitals, *J. Chem. Phys.* **87**, 8, pp. 4696–4699.

Koga, T. and Matsuhashi, T. (1988). One-electron diatomics in momentum space V. nonvariational LCAO approach, *J. Chem. Phys.* **89**, p. 983.

Nakashima, H., Hijikata, Y. and Nakatsuji, H. (2008). Solving the electron and electron-nuclear Schrödinger equation for the excited states of helium atom with the free iterative-complement-interaction method, *J. Chem. Phys.* **128**, p. 154108.

Ralchenko, Y., Kramida, A., Reader, J. and Team, N. A. (2008). NIST Atomic Spectra Database, `http://physics.nist.gov/asd`.

Rotenberg, M. (1970). Theory and application of Sturmian functions, *Adv. At. Mol. Phys.* **6**, pp. 233–68.

Sherstyuk, A. (1983). Sturmian expansions in the many-fermion problem, *Teor. Mat. Fiz.* **56**, 2, pp. 272–287.

Shull, H. and Löwdin, P.-O. (1959). Superposition of configurations and natural spin-orbitals. applications to the He problem, *J. Chem. Phys.* **30**, p. 617.

Szabo, A. and Ostlund, N. (1996). *Modern Quantum Chemistry* (Dover, New York), ISBN 0-486-69186-1.

T., S. and C.E., W. (1965). Molecular orbitals in momentum space, *Proc. Roy. Soc. A* **286**, p. 376.

Wolniewicz, L. (1965). Variational treatment of the HeH^+ ion and the beta-decay in HT, *J. Chem. Phys.* **43**, p. 1087.

Can Coulomb Sturmians be used as a basis for N-electron molecular calculations?

Published in J. Phys. Chem. A: Avery and Avery [2009].

Can Coulomb Sturmians Be Used as a Basis for *N*-Electron Molecular Calculations?

John Avery*,‡ and James Avery*,§

Department of Chemistry, University of Copenhagen, and Department of Computer Science, University of Copenhagen

Received: May 1, 2009; Revised Manuscript Received: August 22, 2009

A method is proposed for using isoenergetic configurations formed from many-center Coulomb Sturmians as a basis for calculations on *N*-electron molecules. Such configurations are solutions to an approximate *N*-electron Schrödinger equation with a weighted potential, and they are thus closely analogous to the Goscinskian configurations that we have used previously to study atomic spectra. We show that when the method is applied to diatomic molecules, all of the relevant integrals are pure functions of the parameter $s = kR$, and therefore they can be evaluated once and for all and stored.

Introduction

Since the 1959 paper by Shull and Löwdin,[1] Coulomb Sturmians have found many successful applications in atomic physics.[2−12] Their completeness combined with good convergence properties make them suitable for representing orbitals of electrons in screened Coulomb potentials.

The importance of Coulomb Sturmians in atomic physics suggests that it would be desirable to extend their use to molecular orbital theory. The momentum-space application of many-center Coulomb Sturmians to molecular calculations was pioneered by Shibuya and Wulfman,[13] Aquilanti and co-workers in Perugia, Italy,[14−19] and by Koga's group in Muroan, Japan.[20,21] Aquilanti and his group have been especially known for their deep mathematical investigations of the relationships between the properties of Sturmians and hyperangular momentum theory.

Molecular calculations with Coulomb Sturmians have been confined to one-electron problems. However, in the present paper, we propose a method for using isoenergetic *N*-electron configurations based on Coulomb Sturmians to treat many-center many-electron problems. We show that when this method is applied to diatomic molecules, all of the relevant integrals are pure functions of the parameter $s = kR$, where k is the exponent common to the one-electron basis set and R is the internuclear distance. Closed-form expressions for the integrals can be calculated once and for all, with s and the nuclear charges Z_a and Z_b left as parameters. We believe this to give molecular orbitals based on Coulomb Sturmians a strong advantage compared to other exponential type orbitals.

Coulomb Sturmians on a Single Center

Coulomb Sturmians have the form

$$\chi_\mu(\mathbf{x}_j) = R_{n,l}(t_j)\, Y_{l,m}(\theta_j,\phi_j) \tag{1}$$

where $t_j \equiv kr_j$ and

$$\mu \equiv (n, l, m) \tag{2}$$

and where the functions $\chi_\mu(\mathbf{x}_j)$ satisfy

$$\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} - \frac{nk}{r_j}\right]\chi_\mu(\mathbf{x}_j) = 0 \tag{3}$$

They obey the potential-weighted orthogonality relation:

$$\int \mathrm{d}^3x_j\, \chi_{\mu'}^*(\mathbf{x}_j)\frac{1}{r_j}\chi_\mu(\mathbf{x}_j) = \frac{k}{n}\delta_{\mu'\mu} \tag{4}$$

From eqs 3 and 4, it follows that

$$\int \mathrm{d}^3x_j\, \chi_{\mu'}^*(\mathbf{x}_j)\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right]\chi_\mu(\mathbf{x}_j) = k^2\delta_{\mu'\mu} \tag{5}$$

Many-Center Coulomb Sturmians in Molecular Calculations

To use a many-center Coulomb Sturmian basis set to treat molecules, we introduce the notation

$$\chi_\tau(\mathbf{x}) \equiv \chi_\mu(\mathbf{x} - \mathbf{X}_a) \tag{6}$$

for an orbital centered on nucleus a, where

$$\tau \equiv (a, l, m, n) \tag{7}$$

In a molecule, the one-electron nuclear attraction potential has the form

$$v(\mathbf{x}_j) = -\sum_a \frac{Z_a}{|\mathbf{x}_j - \mathbf{X}_a|} \tag{8}$$

We can build up solutions to the one-electron equation

† Part of the "Vincenzo Aquilanti Festschrift".

* Corresponding authors. E-mail: avery.john.s@gmail.com; avery@diku.dk.
‡ Department of Chemistry.
§ Department of Computer Science.

10.1021/jp9040502
Published on Web 10/06/2009

$$\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(\mathbf{x}_j)\right]\varphi_\zeta(\mathbf{x}_j) = 0 \tag{9}$$

from superpositions of many-center Coulomb Sturmians:

$$\varphi_\zeta(\mathbf{x}_j) = \sum_\tau \chi_\tau(\mathbf{x}_j) C_{\tau\zeta} \tag{10}$$

Thus we have

$$\sum_\tau \left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(\mathbf{x}_j)\right]\chi_\tau(\mathbf{x}_j) C_{\tau\zeta} = 0 \tag{11}$$

Taking the scalar product with a conjugate Coulomb Sturmian yields the 1-electron secular equations:

$$\sum_\tau \int \mathrm{d}^3x_j\, \chi_{\tau'}^*(\mathbf{x}_j)\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(\mathbf{x}_j)\right]\chi_\tau(\mathbf{x}_j) C_{\tau\zeta} = 0 \tag{12}$$

If we let

$$\mathscr{S}_{\tau'\tau} \equiv \frac{1}{k^2}\int \mathrm{d}^3x_j\, \chi_{\tau'}^*(\mathbf{x}_j)\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right]\chi_\tau(\mathbf{x}_j) \tag{13}$$

and

$$\mathscr{W}'_{\tau'\tau} \equiv -\frac{1}{k}\int \mathrm{d}^3x_j\, \chi_{\tau'}^*(\mathbf{x}_j)\, v(\mathbf{x}_j)\, \chi_\tau(\mathbf{x}_j) \tag{14}$$

then the one-electron secular equations take the form

$$\sum_\tau \left[\mathscr{W}'_{\tau'\tau} - \frac{k}{\beta_\nu}\mathscr{S}_{\tau'\tau}\right]C_{\tau\zeta} = 0 \tag{15}$$

The integrals $\mathscr{S}_{\tau'\tau}$ are the well-studied Shibuya−Wulfman integrals, which can be generated using a variety of algorithms. It can be shown that $\mathscr{W}'_{\tau'\tau}$ and $\mathscr{S}_{\tau'\tau}$ are related through the sum rule:[20,21]

$$\mathscr{W}'_{\tau'\tau''} = \sum_\tau \mathscr{S}_{\tau'\tau}\frac{Z_a}{n}\mathscr{S}_{\tau\tau''} \tag{16}$$

In eq 16, $\tau = (a, \mu)$ runs over all the atoms a in the molecule and the full Coulomb Sturmian basis set centered at $\mathbf{X}_a$. If the basis set is truncated, the relationship is only approximate. We now introduce the matrix

$$\mathbf{K}_{\tau'\tau} \equiv \sqrt{\frac{Z_{a'}Z_a}{n'n}}S_{\tau'\tau} \tag{17}$$

which we can call the *Koga matrix* to honor the important contributions of Professor Toshikatsu Koga and his co-workers. Then the sum rule can be rewritten as

$$\mathscr{W}'_{\tau'\tau} = \sqrt{\frac{n'n}{Z_{a'}Z_a}}\sum_{\tau''}\mathbf{K}_{\tau'\tau''}\mathbf{K}_{\tau''\tau} \tag{18}$$

Substituting this into the one-electron secular equations we obtain

$$\sum_\tau\left[\sum_{\tau''}\boldsymbol{K}_{\tau'\tau''}\boldsymbol{K}_{\tau''\tau} - \frac{k}{\beta_\nu}\boldsymbol{K}_{\tau'\tau}\right]C_{\tau\zeta} = 0 \tag{19}$$

Now suppose that we have a set of coefficients $C_{\tau\zeta}$ that satisfy

$$\sum_\tau\left[\boldsymbol{K}_{\tau'\tau} - \frac{k}{\beta_\nu}\delta_{\tau'\tau}\right]C_{\tau\zeta} = 0 \tag{20}$$

Then

$$\sum_\tau K_{\tau'\tau}C_{\tau\zeta} = \frac{k}{\beta_\nu}C_{\tau'\zeta} \tag{21}$$

If we substitute this into eq 19 and carry out the sum over τ, we obtain 20. Thus we can see that if the coefficients $C_{\tau\zeta}$ satisfy eq 20, then the one-electron secular eq 15 will also be satisfied, provided that the basis set is complete. Otherwise, solutions to eq 20 are approximate solutions to eq 15. Solution of the secular eq 20 also gives us solutions to eq 9. When the basis set is truncated, the 1-electron secular eqs 15 and 20 give slightly different results. In the case of diatomic molecules, we can evaluate $\mathscr{W}$directly without using eq 16 or 18, as we will show in eqs 59−62, and so we can choose between eqs 15 and 20. Because of overcompleteness, eq 15 breaks down numerically for small internuclear distances R but gives superior results to eq 20 for large separations. In contrast, eq 20 does not suffer from overcompleteness and becomes progressively more accurate as R approaches zero. Thus the two forms of the 1-electron secular equations complement each other. Calculated energies using eqs 15 and 20 in the appropriate ranges are shown in Tables 1 and 2.

Molecular Calculations Using the Isoenergetic Configurations $\mathbf{\Phi}_\nu(\mathbf{x})$

We now introduce the N-electron configurations which are Slater determinants of the form

$$\Phi_\nu(\mathbf{x}) = |\varphi_{\zeta_1}\varphi_{\zeta_2}...\varphi_{\zeta_N}| \tag{22}$$

Since the individual molecular orbitals satisfy 9, the configurations $\Phi_\nu(\mathbf{x})$ are solutions to the separable N-electron equation:

$$\sum_{j=1}^N\left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} + \beta_\nu v(\mathbf{x_j})\right]\Phi_\nu(\mathbf{x}) = 0 \tag{23}$$

which can also be written in the form

$$\left[\sum_{j=1}^N\left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right) + \beta_\nu V_0(\mathbf{x})\right]\Phi_\nu(\mathbf{x}) = 0 \tag{24}$$

where

$$V_0(\mathbf{x}) = \sum_{j=1}^{N} v(\mathbf{x}_j) \tag{25}$$

We would like to use these configurations to build up solutions to the *N*-electron Schrödinger equation

$$\left[\sum_{j=1}^{N}\left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right) + V(\mathbf{x})\right]\Psi_\kappa(\mathbf{x}) = 0 \tag{26}$$

with

$$V(\mathbf{x}) = \sum_{j=1}^{N} v(\mathbf{x}_j) + \sum_{i>j}^{N}\sum_{j=1}^{N}\frac{1}{r_{ij}} \tag{27}$$

and with

$$E_\kappa = -\sum_{j=1}^{N}\frac{k^2}{2} = -\frac{Nk^2}{2} \tag{28}$$

Thus we write

$$\Psi_\kappa(\mathbf{x}) = \sum_\nu \Phi_\nu(\mathbf{x}) B_{\nu\kappa} \tag{29}$$

Substituting this into the *N*-electron Schrödinger equation, and taking the scalar product with a conjugate configuration, we obtain the secular equations:

$$\sum_\nu \int \mathrm{d}x\, \Phi_{\nu'}^*(\mathbf{x})\left[\sum_{j=1}^{N}\left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right) + V(\mathbf{x})\right]\Phi_\nu(\mathbf{x})B_{\nu\kappa} = 0 \tag{30}$$

We now introduce a *k*-independent matrix representing the total potential based on the configurations $\Phi_\nu(\mathbf{x})$:

$$T_{\nu'\nu}^{(N)} \equiv -\frac{1}{k}\int \mathrm{d}x\, \Phi_{\nu'}^*(\mathbf{x})\, V(\mathbf{x})\, \Phi_\nu(\mathbf{x}) \tag{31}$$

and another *k*-independent matrix

$$\mathscr{G}_{\nu'\nu}^{(N)} \equiv \frac{1}{k^2}\int \mathrm{d}x\, \Phi_{\nu'}^*(\mathbf{x}) \sum_{j=1}^{N}\left(-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2}\right)\Phi_\nu(\mathbf{x}) \tag{32}$$

In terms of these matrices, the secular equations become

$$\sum_\nu [\mathbf{T}_{\nu'\nu}^{(N)} - k\mathscr{G}_{\nu'\nu}^{(N)}]\mathbf{B}_{\nu\kappa} = 0 \tag{33}$$

Solving eq 33, we obtain *k* for each state κ and thus the energy $E_\kappa = -Nk^2/2$. For a given state κ, the value of *k* then determines the weighting factors β_{ν_1}, β_{ν_2}, ... needed to make each configuration Φ_{ν_1}, Φ_{ν_2}, ... correspond to the same energy E_κ.

Solving the Secular Equations

The matrix $T_{\nu'\nu}^{(N)}$ can be constructed from the 1-electron components using the generalized Slater−Condon rules.[12] From eq 10 we have

$$\tilde{m}_{\zeta'\zeta} \equiv \int \mathrm{d}^3x_j\, \varphi_{\zeta'}^*(\mathbf{x}_j)\, \varphi_\zeta(\mathbf{x}_j) = \sum_{\tau'}\sum_{\tau} C_{\tau'\zeta'}^* m_{\tau'\tau} C_{\tau\zeta} \tag{34}$$

where

$$m_{\tau'\tau} \equiv \int \mathrm{d}^3x_j\, \chi_{\tau'}^*(\mathbf{x}_j)\, \chi_\tau(\mathbf{x}_j) \tag{35}$$

are displaced Sturmian overlap integrals. Similarly

$$\tilde{v}_{\zeta'\zeta} \equiv \int \mathrm{d}^3x_j \varphi_{\zeta'}^*(\mathbf{x}_j) v(\mathbf{x}_j) \varphi_\zeta(\mathbf{x}_j) = -k\sum_{\tau'}\sum_{\tau} C_{\tau'\zeta'}^* \mathscr{M}_{\tau'\tau} C_{\tau\zeta} \tag{36}$$

and also

$$\tilde{\mathscr{G}}_{\zeta'\zeta} \equiv \sum_{\tau'}\sum_{\tau} C_{\tau'\zeta'}^* \mathscr{G}_{\tau'\tau} C_{\tau\zeta} = \frac{\beta_\nu}{k^2}\tilde{v}_{\zeta'\zeta} \tag{37}$$

To build the *N*-electron matrices $T_{\nu'\nu}^{(N)}$ and $\mathscr{G}_{\nu'\nu}^{(N)}$ and solve eq 33, we must first obtain the coefficients $C_{\tau\zeta}$ by solving eq 15 or 20. In the case of diatomic molecules, we begin by picking a value of the parameter $s = kR$, where *R* is the interatomic distance and *k* is the exponent of the Coulomb Sturmian basis set. Neither *R* nor *k* is known at this point, but only their product *s*. As we shall see below, for the diatomic case, all of the integrals involved in eqs 15 and 20 are pure functions of *s*. Having chosen *s*, we can thus solve the one-electron secular equations and obtain the coefficients $C_{\tau\zeta}$ and the spectrum of ratios k/β_ν. We are then able to solve eq 33, which gives us a spectrum of *k*-values, and thus energies $-Nk^2/2$, and the eigenvectors $\mathbf{B}_{\nu\kappa}$. From a *k*-value, we also get the unscaled distance $R = s/k$. We repeat the procedure for a range of *s*-values and interpolate to find the solutions as functions of *R*.

In the case of polyatomic molecules, one can choose a set of angles between the nuclei; these are left fixed under scaling of the coordinate system. The procedure is then similar to that described for the diatomic case.

Evaluation of Shibuya−Wulfman and Sturmian Overlap Integrals

The two-center Sturmian overlap integrals can conveniently be evaluated in momentum space, using the relationships[12]

$$m_{\tau'\tau} \equiv \int \mathrm{d}^3x_j\, \chi_{\tau'}^*(\mathbf{x}_j)\, \chi_\zeta(\mathbf{x}_j) = \int \mathrm{d}^3p\, \chi_{\tau'}^{t*}(\mathbf{p})\, \chi_\tau^t(\mathbf{p}) \tag{38}$$

and

$$\chi_\tau^t(\mathbf{p}) = M(p)\, Y_{n-1,l,m}(\mathbf{u})\mathrm{e}^{-\mathrm{i}\mathbf{p}\cdot\mathbf{X}_a} \tag{39}$$

Here

$$M(p) \equiv \frac{4k^{5/2}}{(k^2 + p^2)^2} \tag{40}$$

while $Y_\mu(\mathbf{u})$ is a 4-dimensional hyperspherical harmonic. In eq 39, $\mathbf{u}$ is a unit vector that defines Fock's projection of momentum space onto the surface of a 4-dimensional hypersphere.[22]

$$\mathbf{u} = (u_1, u_2, u_3, u_4) = \left(\frac{2kp_1}{k^2 + p^2}, \frac{2kp_2}{k^2 + p^2}, \frac{2kp_3}{k^2 + p^2}, \frac{k^2 - p^2}{k^2 + p^2}\right) \tag{41}$$

Substituting eq 39 into eq 38, we obtain

$$m_{\tau'\tau} = \int \mathrm{d}^3p \, \frac{16k^5}{(k^2 + p^2)^4} Y^*_{n'-1,l',m'}(\mathbf{u}) \, Y_{n-1,l,m}(\mathbf{u}) \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} \tag{42}$$

where $\mathbf{R} \equiv \mathbf{X}_{a'} - \mathbf{X}_a$. The integral can be evaluated directly using the expansion of a plane wave in terms of spherical Bessel functions and Legendre polynomials:

$$\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} = \sum_{l=0}^{\infty} i^l (2l + 1) \, j_l(pR) \, P_l(\hat{\mathbf{p}}\cdot\hat{\mathbf{R}}) \tag{43}$$

For example, $Y_{0,0,0} = 1/(2\pi^2)^{1/2}$, and therefore

$$m_{(a',1,0,0),(a,1,0,0)} = \frac{2}{\pi}\int_0^\infty dp \frac{16k^5p^2}{(k^2 + p^2)^4} j_0(pR) = e^{-s}\left(1 + s + \frac{s^2}{3}\right) \tag{44}$$

where $s \equiv kR$. Alternatively, we can convert the integral in eq 42 into a hyperangular integral using the relationship[12]

$$\mathrm{d}^3p = \mathrm{d}\Omega\left(\frac{k^2 + p^2}{2k}\right)^3 \tag{45}$$

Then eq 42 becomes

$$m_{\tau'\tau} = \int \mathrm{d}\Omega \, \frac{2k^2}{(k^2 + p^2)} Y^*_{n'-1,l',m'}(\mathbf{u}) \, Y_{n-1,l,m} \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} \tag{46}$$

or

$$m_{\tau'\tau} = \int \mathrm{d}\Omega \, (1 + u_4) Y^*_{n'-1,l',m'}(\mathbf{u}) \, Y_{n-1,l,m} \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} \tag{47}$$

Similarly, the Shibuya−Wulfman integrals can be expressed in the form[13]

$$\mathscr{S}_{\tau'\tau} = \int \mathrm{d}\Omega \, Y^*_{n'-1,l',m'}(\mathbf{u}) \, Y_{n-1,l,m} \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} \tag{48}$$

We then make use of the relationship[9]

$$\int \mathrm{d}\Omega \, Y_{n-1,l,m} \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} = (2\pi)^{3/2} f_{n,l}(s) \, Y_{l,m}(\hat{\mathbf{s}}) \tag{49}$$

where $Y_{l,m}$ is an ordinary 3-dimensional spherical harmonic and where

$$\mathbf{s} = \{s_x, s_y, s_z\} \equiv k\mathbf{R} \tag{50}$$

The function $f_{n,l}(s)$ is defined by

$$k^{3/2} f_{n,l}(s) \equiv R_{n,l}(s) - \frac{1}{2}\sqrt{\frac{(n - l)(n + l + 1)}{n(n + 1)}} R_{n+1,l}(s) - \frac{1}{2}\sqrt{\frac{(n + l)(n - l - 1)}{n(n - 1)}} R_{n-1,l}(s) \tag{51}$$

where $R_{n,l}$ is the radial function of the Coulomb Sturmians given in eq 1, and where

$$R_{n-1,l}(s) \equiv 0 \quad \text{if} \quad l > n - 1 \tag{52}$$

Similarly, the integral

$$m_{(a',1,0,0),(a,n,l,m)} = \frac{1}{\sqrt{2}\pi} \int \mathrm{d}^3p \, \frac{16k^5}{(k^2 + p^2)^4} Y_{n-1,l,m}(\mathbf{u}) \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} \tag{53}$$

can be written in the form

$$m_{(a',1,0,0),(a,n,l,m)} = \sqrt{4\pi} g_{n,l}(s) \, \mathbf{Y}_{l,m}(\hat{\mathbf{s}}) \tag{54}$$

For other values of τ', $m_{\tau'\tau}$ can be found by resolving $Y^*_{n'-1,l',m'}(\mathbf{u}) \, Y_{n-1,l,m}(\mathbf{u})$ into a sum of 4-dimensional hyperspherical harmonics. The function $g_{n,l}(s)$ is given by

$$g_{n,l}(s) \equiv f_{n,l}(s) - \frac{1}{2}\sqrt{\frac{(n - l)(n + l + 1)}{n(n + 1)}} f_{n+1,l}(s) - \frac{1}{2}\sqrt{\frac{(n + l)(n - l - 1)}{n(n - 1)}} f_{n-1,l}(s) \tag{55}$$

where we define

$$f_{n-1,l}(s) \equiv 0 \quad \text{if} \quad l > n - 1 \tag{56}$$

The coupling coefficients that appear here are the ones needed for resolving $(1 + u_4)Y_{n-1,l,m}$ into a sum of 4-dimensional hyperspherical harmonics. The problem of evaluating both the Shibuya−Wulfman integrals and the two-center Coulomb Sturmian overlap integrals thus reduces to the problem of resolving the product $Y^*_{n'-1,l',m'}Y_{n-1,l,m}$ into a sum of 4-dimensional hyperspherical harmonics. This can be done either using 4-dimensional Wigner coefficients, a method pioneered by Prof. Vincenzo Aquilanti and his co-workers,[14,15,18,19,23] or alternatively, using the harmonic projection methods studied by us.[3,6,7,12] For example, since $Y_{0,0,0}(\mathbf{u}) = 1/(2\pi^2)^{1/2}$ we have

$$\mathcal{S}_{(a',1,0,0),(a,n,l,m)} = \frac{1}{\sqrt{2\pi^2}} \int d\Omega\, Y_{n-1,l,m}(\mathbf{u}) e^{i\mathbf{p}\cdot\mathbf{R}} = \sqrt{4\pi} f_{n,l}(s)\, Y_{l,m}(\hat{\mathbf{s}}) \quad (57)$$

Evaluation of the Nuclear Attraction Integrals, $\tau'\tau$

From the definition of the nuclear attraction integrals (14) and from (8), we have

$$\mathcal{W}_{\tau'\tau} \equiv -\frac{1}{k} \int d^3x_j\, \chi^*_{\tau'}(\mathbf{x}_j)\, v(\mathbf{x}_j)\, \chi_\tau(\mathbf{x}_j) = \frac{1}{k} \int d^3x_j\, \chi^*_{\tau'}(\mathbf{x}_j) \sum_a \frac{Z_a}{|\mathbf{x}_j - \mathbf{X}_a|} \chi_\tau(\mathbf{x}_j) \quad (58)$$

For diatomic molecules, this becomes

$$\mathcal{W}_{\tau'\tau} = \frac{1}{k} \int d^3x_j\, \chi^*_{\tau'}(\mathbf{x}_j) \left[\frac{Z_1}{|\mathbf{x}_j - \mathbf{X}_1|} + \frac{Z_2}{|\mathbf{x}_j - \mathbf{X}_2|} \right] \chi_\tau(\mathbf{x}_j) \quad (59)$$

For the case of diatomic molecules, we are thus faced with three types of integrals:

$$I^{(1)}_{\mu'\mu} = \frac{1}{k} \int d^3x_j\, \chi^*_{\mu'}(\mathbf{x}_j) \frac{1}{|\mathbf{x}_j - \mathbf{X}|} \chi_\mu(\mathbf{x}_j) \quad (60)$$

$$I^{(2)}_{\mu'\mu} = \frac{1}{k} \int d^3x_j\, \chi^*_{\mu'}(\mathbf{x}_j) \frac{1}{|\mathbf{x}_j - \mathbf{X}|} \chi_\mu(\mathbf{x}_j - \mathbf{X}) = \frac{1}{k} \int d^3x_j\, \chi^*_{\mu'}(\mathbf{x}_j) \left[-\frac{1}{2}\nabla_j^2 + \frac{k^2}{2} \right] \chi_\mu(\mathbf{x}_j - \mathbf{X}) \quad (61)$$

and

$$I^{(3)}_{\mu'\mu} = \frac{1}{k} \int d^3\, x_j \chi^*_{\mu'}(\mathbf{x}_j) \frac{1}{|\mathbf{x}_j|} \chi_\mu(\mathbf{x}_j) = \frac{\delta_{\mu'\mu}}{n} \quad (62)$$

Integrals of the type $I^{(1)}_{\mu'\mu}$ can be evaluated by expanding $1/|\mathbf{x}_j - \mathbf{X}|$ in terms of Legendre polynomials, or alternatively by introducing the Fourier integral representation of $1/|\mathbf{x}_j - \mathbf{X}|$, while integrals of the type $I^{(2)}_{\mu'\mu}$ can be expressed in terms of Shibuya−Wulfman integrals. Integrals of the type $I^{(3)}_{\mu'\mu}$ are easily evaluated using eq 5.

Two-Center Interelectron Repulsion Integrals for Coulomb Sturmians

If $\rho_1(\mathbf{x} - \mathbf{X}_a)$ and $\rho_2(\mathbf{x}' - \mathbf{X}_{a'})$ are two electron density distributions, centered respectively on nuclei at the positions $\mathbf{X}_a$ and $\mathbf{X}_{a'}$, the interelectron repulsion between them is given by the integral

$$J = \int d^3x \int d^3x'\, \rho_1(\mathbf{x} - \mathbf{X}_a) \frac{1}{|\mathbf{x} - \mathbf{x}'|} \rho_2(\mathbf{x}' - \mathbf{X}_{a'}) \quad (63)$$

If we introduce the Fourier transform representation

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{2\pi^2} \int d^3p\, \frac{1}{p^2} e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{x}')} \quad (64)$$

we can rewrite J in the form

$$J = 4\pi \int d^3p\, \frac{1}{p^2} \rho^t_1(\mathbf{p})\, \rho^t_2(-\mathbf{p}) e^{i\mathbf{p}\cdot\mathbf{R}} \quad (65)$$

where $\mathbf{R} = \mathbf{X}_{a'} - \mathbf{X}_a$ and

$$\rho^t_j(\mathbf{p}) = \frac{1}{(2\pi)^{3/2}} \int d^3x\, \rho_j(\mathbf{x}) e^{-i\mathbf{p}\cdot\mathbf{x}} \qquad j = 1, 2 \quad (66)$$

The Fourier transforms of products of Coulomb Sturmians are easy to obtain. For example,

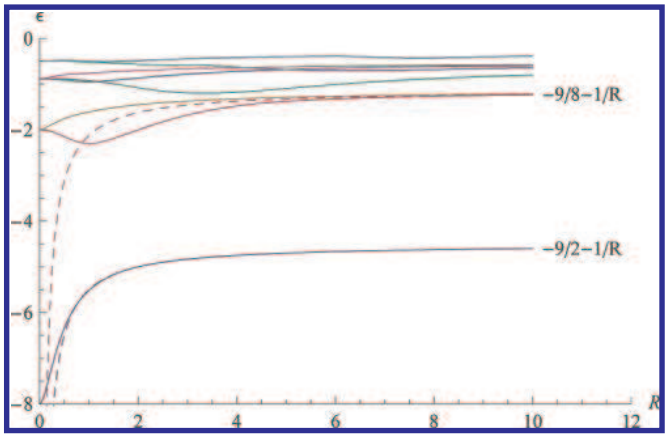


Figure 1. Electronic energies of LiH^{3+} obtained as functions of R from eqs 15 and 20 with $\beta_\nu = 1$ and $\varepsilon = -k^2/2$. The z-axis was taken in the direction of the internuclear separation, and only basis functions with $m = 0$ were used. The dotted lines show the asymptotic behaviors $-9/2 - 1/R$ and $-9/8 - 1/R$ of the lowest three energies (see Table 2). As the LiH^{3+} ion dissociates, the single electron of the system follows the Li nucleus. The electronic energy of the ion approaches that of Li^{2+}, $-9/2$ hartree, plus the energy of the electron in the attractive potential of the distant H^+ ion, $-1/R$.

TABLE 1: Energies as Functions of Internuclear Distances for the Lowest Orbitals of H_2^+, Calculated Using a Basis with 10 Coulomb Sturmians on Each Nucleus (Atomic Units Used Throughout)

	energy		
R	σ_g 1s	σ_g 2s	σ^*_u 2s
0	−2.00000	−0.500000	−0.500000
0.1	−1.97822	−0.500613	−0.496003
0.2	−1.92846	−0.502489	−0.48914
0.4	−1.80006	−0.509921	−0.470570
0.6	−1.67030	−0.522165	−0.452714
0.8	−1.55305	−0.538910	−0.437248
1	−1.45032	−0.559748	−0.423120
2	−1.10220	−0.667529	−0.360727
3	−0.910878	−0.701418	−0.318789
4	−0.796074	−0.695551	−0.289563
5	−0.724413	−0.677292	−0.265364
6	−0.678631	−0.657311	−0.247421
8	−0.627569	−0.623606	−0.221687
10	−0.600578	−0.599901	−0.204651
$R \to \infty$	$-1/2 - 1/R$	$-1/2 - 1/R$	$-1/8 - 1/R$

If $\rho_{(1s)^2}(\mathbf{x}) = [\chi_{1,0,0}(\mathbf{x})]^2$ then $\rho^t_{(1s)^2}(\mathbf{p}) = \frac{4\sqrt{2}k^4}{(4k^2 + p^2)^2\pi^{3/2}}$ (67)

The integrals in eq 65 can be evaluated analytically. For example,

$$J_{\langle(1s)^2|(1s)^2\rangle} = (4\pi)^2 \int_0^\infty \mathrm{d}p \left[\frac{4\sqrt{2}k^4}{(4k^2 + p^2)^2\pi^{3/2}}\right]^2 j_0(pR) = \frac{24 + \mathrm{e}^{-2kR}(-24 - kR(33 + 2kR(9 + 2kR)))}{24R} \quad (68)$$

This is the Coulomb integral representing the interelectron repulsion between a $(1s)^2$ Coulomb Sturmian charge density located at the point $\mathbf{X}_a$ and a $(1s)^2$ density located at $\mathbf{X}_{a'}$. When $R \rightarrow 0$, the integral approaches $5k/8$, while for large R, it approaches $1/R$. Notice that if we divide it by k, the integral is a pure function of $s = kR$, and this is true of all integrals of this type representing interactions between products of Coulomb Sturmians. Similarly, we find that

$$\frac{1}{k}J_{\langle(1s)^2|(2s)^2\rangle} = \frac{1}{s} + \frac{\mathrm{e}^{-2s}}{240s}[-240 - s(390 + s(300 + s(145 + 2s(25 + 6s))))] \quad (69)$$

and so on. To systematize integrals of this type, we note that

$$\mathscr{I}_{n,l}\left(\frac{p}{2k}\right) \equiv \int_0^\infty \mathrm{d}t\, t^n \mathrm{e}^{-2t} j_l(pt/k) = \left(\frac{p}{2k}\right)^l \frac{(n + l)!\,{}_2F_1[(a|b|c - (p/2k)^2}{(2l + 1)!!2^{n+1}[1 + (p/2k)^2]^n} \quad (70)$$

where

$$a \equiv (l - n + 1)2$$
$$b \equiv (l - n + 1)2$$
$$c \equiv l + 3/2$$

and that the Fourier transform of the partial density $\rho_{jlm}(\mathbf{x}) = 4k^3(kr)^j \mathrm{e}^{-2kr} Y_{l,m}(\hat{\mathbf{x}})$ is therefore given by

TABLE 2: Energies as Functions of Internuclear Distances for the Lowest Orbitals of the LiH^{3+} Ion (Figure 1)

	energy		
R	σ_g 1s	σ_g 2s	σ_u^* 2s
0	−8.00000	−2.00000	−2.00000
0.1	−7.77863	−2.00743	−1.96276
0.2	−7.38734	−2.02913	−1.91086
0.4	−6.64066	−2.10542	−1.80342
0.6	−6.10336	−2.19682	−1.71865
0.8	−5.75353	−2.27294	−1.66012
1	−5.52094	−2.30570	−1.61374
2	−5.00204	−2.00153	−1.45288
3	−4.83370	−1.67811	−1.37216
4	−4.75011	−1.48006	−1.32336
5	−4.70005	−1.37882	−1.29064
6	−4.66669	−1.32515	−1.26717
8	−4.62501	−1.26748	−1.23576
10	−4.60000	−1.23586	−1.21571
$R \rightarrow \infty$	$-9/2 - 1/R$	$-9/8 - 1/R$	$-9/8 - 1/R$

$$\rho^t_{jlm}(\mathbf{p}) = \frac{1}{(2\pi)^{3/2}} \int \mathrm{d}^3x\, \mathrm{e}^{-\mathrm{i}\mathbf{p}\cdot\mathbf{r}}[4k^3(kr)^j \mathrm{e}^{-2kr} Y_{l,m}(\hat{\mathbf{x}})] = 4\sqrt{\frac{2}{\pi}}(-\mathrm{i})^l Y_{l,m}(\hat{\mathbf{p}})\mathscr{I}_{j+2,l}\left(\frac{p}{2k}\right) \quad (71)$$

From eq 65 we can see that the Coulomb interaction between two charge densities of this type, located respectively on the centers $\mathbf{X}_a$ and $\mathbf{X}_{a'}$ will be given by

$$\mathbf{J}_{j'l'm',jlm} = 4\pi \int \mathrm{d}^3p\, \frac{1}{p^2} \rho^t_{j'l'm'}(\mathbf{p})\, \rho^t_{jlm}(-\mathbf{p})\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} \quad (72)$$

Substituting eq 10 into eq 11, we have

$$J_{j'l'm',jlm} = 4\pi\left(4\sqrt{\frac{2}{\pi}}\right)^2 (-\mathrm{i})^{l'+l} \int_0^\infty \mathrm{d}p\, \mathscr{I}_{j'+2,l'}\left(\frac{p}{2k}\right)\mathscr{I}_{j+2,l}\left(\frac{p}{2k}\right) \times \int \mathrm{d}\Omega_p\, Y_{l',m'}(\mathbf{p})\, Y_{l,m}(-\hat{\mathbf{p}})\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} \quad (73)$$

To perform the momentum-space angular integration, we expand the plane wave in terms of Legendre polynomials and spherical Bessel functions:

$$\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} = \sum_{l''=0}^\infty i^{l''}(2l'' + 1) j_{l''}(pR)\, P_{l''}(\hat{\mathbf{p}}\cdot\hat{\mathbf{R}}) \quad (74)$$

We then obtain

$$\int \mathrm{d}\Omega_p\, Y_{l',m'}(\hat{\mathbf{p}})\, Y_{l,m}(-\hat{\mathbf{p}})\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} = \sum_{l''=0}^\infty i^{l''}(2l'' + 1) j_{l''}(pR) \int \mathrm{d}\Omega_p\, Y_{l',m'}(\hat{\mathbf{p}})\, Y_{l,m}(-\hat{\mathbf{p}})\, P_{l''}(\hat{\mathbf{p}}\cdot\hat{\mathbf{R}}) \quad (75)$$

which can be rewritten in the form[12]

$$\int \mathrm{d}\Omega_p\, Y_{l',m'}(\hat{\mathbf{p}})\, Y_{l,m}(-\hat{\mathbf{p}})\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{R}} = 4\pi \sum_{l''=0}^\infty i^{l''} j_{l''}(pR)\, O_{l''}[Y_{l',m'}(\hat{\mathbf{R}})\, Y_{l,m}(-\hat{\mathbf{R}})] \quad (76)$$

where $O_{l''}$ is an operator that projects out the part of an angular function corresponding to angular momentum l''. This gives us finally the result:

$$J_{j'l'm',jlm} = (4\pi)^2\left(4\sqrt{\frac{2}{\pi}}\right)^2 (-\mathrm{i})^{l'+l} \sum_{l''=0}^\infty i^{l''} O_{l''}[Y_{l',m'}(\hat{\mathbf{R}})\, Y_{l,m}(-\hat{\mathbf{R}})] \times \int_0^\infty \mathrm{d}p\, \mathscr{I}_{j'+2,l'}\left(\frac{p}{2k}\right)\mathscr{I}_{j+2,l}\left(\frac{p}{2k}\right) j_{l''}(pR) \quad (77)$$

The integrals $J_{j'l'm',jlm}$ can be precalculated analytically and stored. When $\hat{\mathbf{R}}$ is in the direction of the z-axis, only the case where $m' = -m$ gives a nonzero result. From the stored pieces, we can assemble the integrals in which we are really interested. We let

$$\rho_1(\mathbf{x}) \equiv \chi^*_{\mu_1}(\hat{\mathbf{x}})\,\chi_{\mu_2}(\hat{\mathbf{x}}) = \sum_{j=l_1+l_2}^{n_1+n_2-2} \sum_{l=|l_1-l_2|}^{l_1+l_2} \rho_{j,l,m_2-m_1}(\mathbf{x}) C^{j,l}_{\mu_1,\mu_2} \tag{78}$$

while

$$\rho_2(\mathbf{x}') \equiv \chi^*_{\mu_3}(\hat{\mathbf{x}}')\,\chi_{\mu_4}(\hat{\mathbf{x}}') = \sum_{j'=l_3+l_4}^{n_3+n_4-2} \sum_{l'=|l_3-l_4|}^{l_3+l_4} \rho_{j',l',m_4-m_3}(\mathbf{x}') C^{j',l'}_{\mu_3,\mu_4} \tag{79}$$

and where we have used the abreviation $\mu_i \equiv (n_i, l_i, m_i)$. Then the following formula assembles the integral in which we are interested from the stored pieces:

$$J_{\mu_1,\mu_2,\mu_3,\mu_4} = \sum_{j=l_1+l_2}^{n_1+n_2-2} \sum_{l=|l_1-l_2|}^{l_1+l_2} \sum_{j'=l_3+l_4}^{n_3+n_4-2} \sum_{l'=|l_3-l_4|}^{l_3+l_4} J_{j,l,m_2-m_1,j',l',m_4-m_3} C^{j,l}_{\mu_1,\mu_2} C^{j',l'}_{\mu_3,\mu_4} \tag{80}$$

We now need to calculate the coefficients $C^{j,l}_{\mu_1,\mu_2}$ and $C^{j',l'}_{\mu_3,\mu_4}$. If we introduce the notation:

$$R_{n_1,l_1}(kr)\,R_{n_2,l_2}(kr) = \sum_{j=l_1+l_2}^{n_1+n_2-2} 4k^3 (kr)^j \mathrm{e}^{-2kr} \alpha^j_{n_1,l_1,n_2,l_2} \tag{81}$$

and

$$Y^*_{l_1,m_1}(\mathbf{x})\,Y_{l_2,m_2}(\hat{\mathbf{x}}) = \sum_{l=|l_1-l_2|}^{l_1+l_2} Y_{l,m_2-m_1}(\hat{\mathbf{x}}) \beta^l_{l_1,m_1,l_2,m_2} \tag{82}$$

then

$$C^{j,l}_{\mu_1,\mu_2} = \alpha^j_{n_1,l_1,n_2,l_2} \beta^l_{l_1,m_1,l_2,m_2} \tag{83}$$

Integrals involving densities formed from two-center products of Coulomb Sturmians are more difficult to evaluate, but they too are pure functions of s when divided by k. To treat these integrals, one can, for example, use the Mulliken approximation. Within the framework of the Mulliken approximation, it is possible, in the case of diatomic molecules, to find coefficients $\tilde{C}^{j,l,1}_{\zeta,\zeta'}(s)$ and $\tilde{C}^{j,l,2}_{\zeta,\zeta'}(s)$ such that

$$\rho_{\zeta,\zeta'}(\mathbf{x}) \equiv \varphi^*_{\zeta}(\mathbf{x})\,\varphi_{\zeta'}(\mathbf{x}) \approx \sum_j \sum_l [\rho_{j,l,m'-m}(\mathbf{x} - \mathbf{X}_1)\tilde{C}^{j,l,1}_{\zeta,\zeta'}(s) + \rho_{j,l,m'-m}(\mathbf{x} - \mathbf{X}_2)\tilde{C}^{j,l,2}_{\zeta,\zeta'}(s)] \tag{84}$$

In other words, it is possible to express the density due to a product of two molecular orbitals directly in terms of the partial densities in a manner analogous to eq 78. This was done in the small example given below.

N-Electron Diatomic Molecules

Let us now specialize to the case of diatomic molecules. Without loss of generality, we can choose the z-axis along the direction of internuclear separation. Then, expressing $\mathbf{s}$ in spherical polar coordinates, we have $\theta = 0$, and the spherical harmonics $Y_{lm}(\hat{\mathbf{s}})$ vanish unless $m = 0$; thus also the ϕ-dependence disappears. From eqs 50 and 51 it follows that the Shibuya−Wulfman integrals $\mathscr{S}_{\tau'\tau}$ are sums of terms, each of whose angular dependence is a spherical harmonic $Y_{lm}(\mathbf{s})$, but as we have just seen, these spherical harmonics will simply be constants. Therefore, the Shibuya−Wulfman integrals, for the diatomic case, are pure functions of $s = kR$. By similar arguments, it is easy to see that also $m_{\tau'\tau}$, $\mathscr{W}'_{\tau'\tau}$ and the interelectron repulsion integrals become pure functions of s.

Precomputed Integrals. It is possible to precalculate closed form expressions for $\mathscr{S}_{\tau'\tau}$, $\mathscr{W}'_{\tau'\tau}$ and $m_{\tau'\tau}$, where $\mathscr{S}$ and m depend only on s, and where $\mathscr{W}'$ depends on s and the nuclear charges Z_a and Z_b.

We have calculated closed form expressions for the matrix elements $\mathscr{S}_{\tau'\tau}$, $\mathscr{W}'_{\tau'\tau}$, and $m_{\tau'\tau}$ for the case of diatomic molecules for all valid (n, l, m), (n', l', m') with $n, n' \leq 5$, for all valid $(n, l, 0)$, $(n', l', 0)$ with $n, n' \leq 7$ and for all valid $(n, 0, 0)$, $(n', 0, 0)$ when $n, n' \leq 20$. From the expressions, we have generated very fast C-code, which is made available through a shared library. The implementation, which will be described in a separate publication, can be downloaded from http://sturmian.kvante.org/diatomic.

It contains both the precomputed integrals and the programs that precalculate them using harmonic projection methods and that generate fast C-code. The work of precalculating the interelectron repulsion integrals is in progress.

Calculations on Diatomic Systems. Using the precomputed integrals discussed above, we can with very little additional effort study any N-electron diatomic molecule, obtaining properties as functions of R. We can ensure that the functions will be defined in the desired range of R-values by choosing a sufficiently large range for s, as is done in step 2 below. The following solves a diatomic problem in the range $[0; R_{\max}]$:

1. Choose nuclear charges Z_a, Z_b and one-electron bases $\{\mu_1, ..., \mu_{M_a}\}$ on center a and $\{\mu_1, ..., \mu_{M_b}\}$ on center b. Choose also the N-electron basis set of configurations $\{\nu_1, ..., \nu_M\}$.
2. Let $k_{\max} = (-2E_0/N)^{1/2}$, where E_0 is some lower bound to the energy; for example, the energy of the noninteracting system. This makes sure that $k_{\max}R_{\max} \geq kR_{\max}$.
3. Choose a number of sample points $S = \{0, s_1, ..., s_m, k_{\max}\cdot R_{\max}\} \subset [0; R_{\max}]$.
4. Now, for each sample point $s \in S$, do the following:
   (a) Build the one-electron matrices $\mathscr{S}(Z_a, Z_b, s)$ and $\mathscr{W}'(Z_a, Z_b, s)$ and solve eq 15 or 20. This gives us the roots k/β_ν and coefficients $C_{\tau\zeta}$ for s.
   (b) Use $C_{\tau\zeta}$ to construct $T^{(N)}_{\nu'\nu}$ and $\mathscr{S}^{(N)}_{\nu'\nu}$, by way of eqs 36, 37, and 84, and solve eq 33, yielding $B_{\nu\kappa}(s)$ and the energies $E_\kappa(s) = -Nk^2(s)/2$. For each state κ of the system, we also obtain a corresponding distance $R_\kappa = s/k(s)$.
5. For the states of interest, construct tables listing R-values and the corresponding properties. Finally, interpolating the tables will give smooth functions of the properties in terms of R, ensured to be defined in the range $[0; R_{\max}]$.

A Simple Example: The HeH $^+$ Ion. To illustrate the method, we performed a rough calculation on the HeH^+ ion, making use of the Mulliken approximation to treat two-center densities. For simplicity, we used only one configuration. Our Coulomb Sturmian basis set was also minimal, and used only

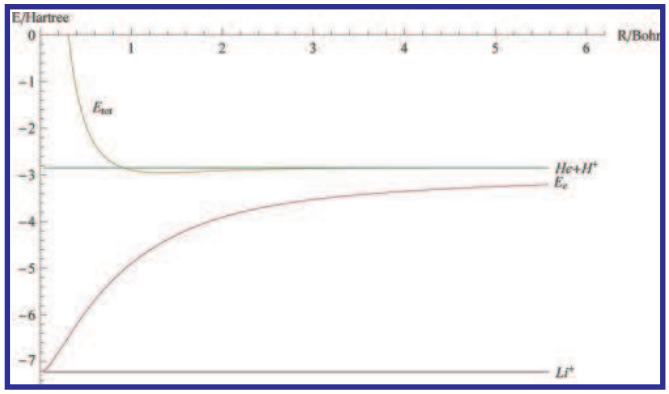


Figure 2. Electronic energy E_e and the total energy E_{tot} of the HeH^+ ion as a function of the internuclear separation $R = s/k$. The calculation was performed with a single configuration using a one-electron basis set consisting of 6 Coulomb Sturmians. For $R \rightarrow 0$, the electronic energy approaches the energy calculated for the Li^+ ion using the Generalized Sturmian Method with a single configuration.[1] In the separated atom region, the total energy approaches that of He when calculated in the same way. Our calculation shows a shallow minimum at $R = 1.35$ Bohrs, which can be compared to the equilibrium bond length of 1.46 Bohrs obtained in a benchmark calculation by Wolniewicz.[24]

six (n, l, m)-values: (1, 0, 0), (2, 0, 0), (2, 1, 0), (3, 0, 0), (3, 1, 0) and (3, 2, 0). We used the single configuration $\Phi_1 = |\varphi_1\varphi_{\bar{1}}|$ where φ_1 and $\varphi_{\bar{1}}$ are the lowest spin-up and spin-down orbitals found by solving eq 15. Figure 2 shows the resulting ground-state energy as a function of the internuclear separation, with and without the internuclear repulsion term; i.e., it shows the electronic energy and the total energy. In the united-atom limit, the electronic energy reduces to $-1849/256 = -0.722266$ hartree. This is the Li^+ energy that is obtained by substituting $Z = 3$ into the formula for the ground-state energies of the 2-electron isoelectronic atomic series, calculated with a single Goscinskian configuration:[1]

$$E = -\frac{1}{2}\left(Z\sqrt{2} - \frac{5}{8\sqrt{2}}\right)^2 \qquad (85)$$

In the separated-atom limit, the total energy approaches the value $-729/256 = -2.84766$ hartree, which is the value obtaifned from eq 85 with $Z = 2$, i.e., the approximate ground-state energy of helium calculated with a single Goscinskian configuration. We can see from this that as HeH^+ dissociates, both electrons follow the helium nucleus. This is why the dissociation of HeH^+ can be described with approximate correctness by a single configuration, which is famously not possible for $\mathbf{H}_2$. Our rough calculation shows a shallow minimum of the total energy at an internuclear separation of 1.34781 Bohrs. This calculated equilibrium bond length is to be compared with the benchmark value,[24] 1.46324 Bohrs. We believe that a more refined calculation, using the present method with a richer basis of Coulomb Sturmians and a larger number of configurations, would lead to closer agreement. With more configurations, we would also obtain the excited states of the system. Greater accuracy could be obtained by avoiding the Mulliken approximation.

Discussion

Two-center Coulomb Sturmians have long been used in 1-electron calculations on diatomic molecules, but let us try to answer the question of our title: Can Coulomb Sturmians be used as basis functions for N-electron molecular calculations? We believe that the method proposed in this paper can make this possible, and that many-center Coulomb Sturmians have some advantages over other ETO basis functions. If we apply the method to diatomic molecules, all of the relevant integrals are pure functions of the parameter $s = kR$ and nuclear charges Z_a and Z_b. The matrices $\mathscr{W}_{\tau'\tau}$, $\mathscr{G}_{\tau'\tau}$, $K_{\tau'\tau}$, and $m_{\tau'\tau}$ can be evaluated once and for all and stored with general nuclear charges and general s. Furthermore, the Fock projection and the theory of hyperspherical harmonics can be used for the efficient evaluation of these integrals. The two-center interelectron repulsion integrals are also pure functions of the parameter $s = kR$, and they too can be evaluated once and for all and stored; however, this work remains to be completed. The isoenergetic N-electron configurations proposed here are closely analogous to the Goscinskian configurations,[2] which we have successfully used to calculate the spectra of atoms and atomic ions.[3−12] Thus the proposed method seems promising to us, and we hope to explore it further.

References and Notes

(1) Shull, H.; Löwdin, P.-O. *J. Chem. Phys.* **1959**, *30*, 617.
(2) Goscinski, O. *Adv. Quantum Chem.* **2003**, *41*, 51−85, Originally unpublished research report, Quantum Chemistry Group, Uppsala University, 1968.
(3) Avery, J. *Hyperspherical Harmonics; Applications in Quantum Theory*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989.
(4) Avery, J. Hyperspherical harmonics; Some properties and applications. In *Conceptual Trends in Quantum Chemistry*; Kryachko, E., Calais, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994.
(5) Avery, J. *J. Math. Chem.* **1997**, *21*, 285.
(6) Avery, J. *J. Math. Chem.* **1998**, *24*, 169.
(7) Avery, J. *Hyperspherical Harmonics and Generalised Sturmians*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000.
(8) Avery, J.; Sauer, S. In *Many-electron Sturmians applied to molecules*; Quantum Systems in Chemistry and Physics, Vol. 1; Hernández-Laguna, A., Maruani, J., McWeeney, R., Wilson, S., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000.
(9) Avery, J. Sturmians. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; Wiley: Chichester, 2003.
(10) Avery, J.; Avery, J.; Aquilanti, V.; Caligiana, A. *Adv. Quantum Chem.* **2004**, *47*, 156–173.
(11) Avery, J.; Avery, J. *J. Phys. Chem. A* **2004**, *41*, 8848–8851.
(12) Avery, J.; Avery, J. *Generalised Sturmians and Atomic Spectra*; World Scientific: Singapore, 2006.
(13) Shibuya, T.; Wulfman, C. *Proc. R. Soc. A* **1965**, *286*, 376.
(14) Aquilanti, V.; Cavalli, S.; Fazio, D. D.; Grossi, G. Hyperangular momentum: Applications to atomic and molecular science. In *New Methods in Quantum Chemistry*; Tsipis, C., Popov, V., Herschbach, D., Avery, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1996.
(15) Aquilanti, V.; Cavalli, S.; Coletti, C.; Grossi, G. *Chem. Phys.* **1996**, *209*, 405.
(16) V. Aquilanti, S. C.; Coletti, C. *Chem. Phys.* **1997**, *214*, 1.
(17) Aquilanti, V.; Caligiana, A. *Chem. Phys. Lett.* **2002**, *366*, 157.
(18) Aquilanti, V.; Caligiana, A. Fundamental World of Quantum Chemistry: A Tribute to the Memory of P. O. Löwdin, I. In *Fundamental World of Quantum Chemistry: A Tribute to the Memory of P. O. Löwdin*; Brändas, E., Kryachko, E., Eds.; Kluwer Academic Publishers: Dordrecht, 2003; p 297.
(19) Caligiana, A. Sturmian Orbitals in Quantum Chemistry. Ph.D. Thesis, University of Perugia, 2003.
(20) Koga, T.; Matsuhashi, T. *J. Chem. Phys.* **1987**, *87*, 4696–4699.
(21) Koga, T.; Matsuhashi, T. *J. Chem. Phys.* **1988**, *89*, 983.
(22) Fock, V. *Kgl. Norske Videnskab. Forh.* **1958**, *31*, 138.
(23) Wen, Z.-Y.; Avery, J. *J. Math. Phys.* **1985**, *26*, 396.
(24) Wolniewicz, L. *J. Chem. Phys.* **1965**, *43*, 1087; DOI: 10.1063/1.1696885.

JP9040502

Atomic core-ionization energies; approximately piecewise-linear and linear relationships

Published in J. Math. Chem.: Avery and Avery [2008].

J Math Chem (2009) 46:164–181
DOI 10.1007/s10910-008-9450-z

ORIGINAL PAPER

Atomic core-ionization energies; approximately piecewise-linear and linear relationships

James Avery · John Avery

Received: 4 June 2008 / Accepted: 5 August 2008 / Published online: 30 August 2008

Abstract In the generalized Sturmian method, solutions to the many-particle Schrödinger equation are built up from isoenergetic sets of solutions to an approximate Schrödinger equation with a weighted potential $\beta_\nu \mathsf{V}_0(\mathbf{x})$. The weighting factors β_ν are chosen in such a way as to make all of the members of the basis set correspond to the energy of the state being represented. In this paper we apply the method to core ionization in atoms and atomic ions, using a basis where $\mathsf{V}_0(\mathbf{x})$ is chosen to be the nuclear attraction potential. We make use of a large-Z approximation, which leads to extremely simple closed-form expressions not only for energies, but also for values of the electronic potential at the nucleus. The method predicts approximately piecewise linear dependence of the core-ionization energies on the number of electrons N for isonuclear series, and an approximately linear dependence of $\Delta E - Z^2/2$ on the nuclear charge Z for isoelectronic series.

Keywords Generalized Sturmians · Large-Z approximation · Atomic structure · Core-ionization · Electronic potential

1 Introduction

Because of their completeness properties, one-electron Sturmian basis sets have long been used in atomic physics. The members of such basis sets are isoenergetic solutions

J. Avery (✉)
Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
e-mail: avery@diku.dk

J. Avery
H. C. Ørsted Institute, University of Copenhagen, Copenhagen, Denmark
e-mail: avery.john.s@gmail.com

to an approximate Schrödinger equation with a weighted potential, the weighting factor being chosen so that all the members of the set correspond to the same energy. For example, Coulomb Sturmian basis sets are square-integrable isoenergetic sets of functions that obey

$$\left[-\frac{1}{2}\nabla^2 - \beta_n \frac{1}{r} - E\right]\phi_{nlm}(\mathbf{x}) = 0 \tag{1}$$

If the weighting factor β_n is chosen to be $\beta_n = kn$ for a positive k, then all the solutions correspond to the energy $E = -k^2/2$. In Eq. 1, and throughout this paper, atomic units are used, so that all energies are in Hartrees and all lengths in Bohrs.

In 1968, Goscinski [19] generalized the Sturmian concept by introducing basis sets that are solutions to an approximate many-particle Schrödinger equation with a weighted potential:

$$\left[-\sum_{j=1}^{N}\frac{1}{2m_j}\nabla_j^2 + \beta_\nu \mathsf{V}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) - E\right]\Phi_\nu(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = 0 \tag{2}$$

the weighting factor β_ν once again being chosen in such a way as to make all of the solutions correspond to the same energy. Basis sets of this kind can be used in a wide variety of problems.

In the present paper, Sect. 2–5 review the generalized Sturmian method as well as the large-Z approximation for atoms and ions that was introduced by us in [17]. These methods have been described in much more detail in our recent book [22]. Other work on Sturmians and generalized Sturmians can be found in references [2–21].

In the remainder of the paper, we apply the generalized Sturmian method in the large-Z approximation to atomic core ionization. The generalized Sturmian method in the large-Z approximation yields strikingly simple expressions for the electronic potential at the nuclei of atoms and atomic ions, as is shown in Sect. 6.

The method also yields extremely simple closed form expressions for the approximate energies of both the ground states and excited states of atoms and atomic ions. In Sect. 7, these closed form expressions are used to derive an approximate linear dependence of the core ionization energies of isoelectronic series on the nuclear charge Z, and an approximate piecewise linear dependence for isonuclear series on the number of electrons N. The expression for the potential at the nucleus gives insight into the piecewise linear core ionization energies of isonuclear series.

The accuracy of the large-Z approximation for few-electron systems is such that even for moderate values of Z, inaccuracies are much smaller than relativistic corrections. An approximate method for making relativistic corrections is introduced in Sect. 8. It is shown that the corrected energies rapidly approach the experimental ones as Z increases.

2 The generalized Sturmian method applied to atoms

In atomic units, the non-relativistic Schrödinger equation for an N-electron atom or atomic ion with nuclear charge Z is given by

$$\left[-\frac{1}{2}\Delta + \mathsf{V}(\mathbf{x}) - E_\kappa\right]\Psi_\kappa(\mathbf{x}) = 0 \tag{3}$$

where Ψ_κ is the κth electronic state with E_κ the corresponding energy,

$$\mathsf{V}(\mathbf{x}) = -\sum_{j=1}^{N}\frac{Z}{r_j} + \sum_{j>i}^{N}\sum_{i=1}^{N}\frac{1}{r_{ij}} \tag{4}$$

and

$$-\frac{1}{2}\Delta \equiv -\frac{1}{2}\sum_{j=1}^{3N}\frac{\partial^2}{\partial x_j^2} \tag{5}$$

The solution is usually built up from a superposition of basis functions

$$\Psi_\kappa(\mathbf{x}) = \sum_\nu \Phi_\nu(\mathbf{x})B_{\nu\kappa} \tag{6}$$

In the generalized Sturmian method, these basis functions are chosen to be isoenergetic solutions to an approximate Schrödinger equation with a weighted potential [2–21]:

$$\left[-\frac{1}{2}\Delta + \beta_\nu \mathsf{V}_0(\mathbf{x}) - E_\kappa\right]\Phi_\nu(\mathbf{x}) = 0 \tag{7}$$

The weighting factors β_ν are especially chosen so that E_κ is the energy of the state to be represented. For few-electron atoms or atomic ions, it is convenient to let $\mathsf{V}_0(\mathbf{x})$ be the electrostatic attraction potential of the nucleus:

$$\mathsf{V}_0(\mathbf{x}) = -\sum_{j=1}^{N}\frac{Z}{r_j} \tag{8}$$

such that

$$\mathsf{V}(\mathbf{x}) = \mathsf{V}_0(\mathbf{x}) + \mathsf{V}'(\mathbf{x}) \;\text{ with }\; \mathsf{V}'(\mathbf{x}) = \sum_{j>i}^{N}\sum_{i=1}^{N}\frac{1}{r_{ij}} \tag{9}$$

With this choice of $\mathsf{V}_0(\mathbf{x})$, the weighting factors β_ν are determined automatically, and Eq. 7 is satisfied by Slater determinants of the form:

$$\Phi_\nu(\mathbf{x}) = |\chi_{\mu_1}\chi_{\mu_2}\ldots\chi_{\mu_N}| \tag{10}$$

where the χ_μ's are hydrogenlike spin-orbitals

$$\begin{aligned}\chi_{nlm,+1/2}(\mathbf{x}_j) &= R_{nl}(r_j)Y_{lm}(\theta_j,\phi_j)\alpha(j)\\ \chi_{nlm,-1/2}(\mathbf{x}_j) &= R_{nl}(r_j)Y_{lm}(\theta_j,\phi_j)\beta(j)\end{aligned} \tag{11}$$

with the weighted charges ([22], Chap. 3):

$$Q_\nu = \beta_\nu Z = \frac{p_\kappa}{\mathcal{R}_\nu} \tag{12}$$

where

$$p_\kappa \equiv \sqrt{-2E_\kappa} \tag{13}$$

and

$$\mathcal{R}_\nu \equiv \sqrt{\frac{1}{n_1^2}+\frac{1}{n_2^2}+\cdots+\frac{1}{n_N^2}} \tag{14}$$

Here $n_1, n_2, \ldots, n_N$ represent the principal quantum numbers of the hydrogenlike spin-orbitals in the configuration Φ_ν. The energy E_κ will then be related to the weighted nuclear charges Q_ν by

$$E_\kappa = -\frac{p_\kappa^2}{2} = -\frac{1}{2}Q_\nu^2\mathcal{R}_\nu{}^2 = -\left(\frac{Q_\nu^2}{2n_1^2}+\frac{Q_\nu^2}{2n_2^2}+\cdots+\frac{Q_\nu^2}{2n_N^2}\right) \tag{15}$$

Each of the hydrogenlike spin-orbitals obeys a one-electron Schrödinger equation of the form:

$$\left[-\frac{1}{2}\nabla_j^2+\frac{Q_\nu^2}{2n^2}-\frac{Q_\nu}{r_j}\right]\chi_\mu(\mathbf{x_j}) = 0 \tag{16}$$

From Eq. 16 it follows that

$$\begin{aligned}\left[-\frac{1}{2}\sum_{j=1}^{N}\nabla_j^2\right]\Phi_\nu(\mathbf{x}) &= \left[-\left(\frac{Q_\nu^2}{2n_1^2}+\frac{Q_\nu^2}{2n_2^2}+\cdots+\frac{Q_\nu^2}{2n_N^2}\right)\right.\\ &\quad\left.+\left(\frac{Q_\nu}{r_1}+\frac{Q_\nu}{r_2}+\cdots+\frac{Q_\nu}{r_N}\right)\right]\Phi_\nu(\mathbf{x})\\ &= \left[E_\kappa-\beta_\nu V_0(\mathbf{x})\right]\Phi_\nu(\mathbf{x})\end{aligned} \tag{17}$$

from which it can be seen that Eq. 7 will indeed be satisfied by the configurations Φ_ν, provided that the effective nuclear charges Q_ν are chosen according to the rule given in Eqs. 12–14. We shall call such a set of isoenergetic solutions to (7) with $V_0(\mathbf{x})$ chosen

to be the nuclear attraction potential a set of "Goscinskian configurations" to honor Prof. Osvaldo Goscinski's important early contributions to the generalized Sturmian method [19].

3 Potential-weighted orthonormality relations

A set of Goscinskian configurations obey potential-weighted orthonormality relations. This can be seen as follows: by rearranging the terms in Eq. 7 we obtain:

$$\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\left[\frac{1}{2}\Delta + E_\kappa\right]\Phi_\nu(\mathbf{x}) = \beta_\nu \int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_\nu(\mathbf{x}) \tag{18}$$

and similarly

$$\int d\tau\ \Phi^*_{\nu}(\mathbf{x})\left[\frac{1}{2}\Delta + E_\kappa\right]\Phi_{\nu'}(\mathbf{x}) = \beta_{\nu'} \int d\tau\ \Phi^*_{\nu}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_{\nu'}(\mathbf{x}) \tag{19}$$

Subtracting the complex conjugate of (19) from (18) and making use of the Hermiticity of the kinetic energy operator, we obtain

$$(\beta_\nu - \beta_{\nu'}) \int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_\nu(\mathbf{x}) = 0 \tag{20}$$

from which it follows that

$$\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_\nu(\mathbf{x}) = 0 \quad \text{if} \quad \beta_{\nu'} \neq \beta_\nu \tag{21}$$

The hydrogenlike spin orbitals are orthonormal:

$$\int d\tau_j\ \chi^*_{\mu'}(\mathbf{x}_j)\chi_\mu(\mathbf{x}_j) = \delta_{\mu'\mu} \tag{22}$$

and they obey the Virial Theorem:

$$-\int d\tau_j\ |\chi_\mu(\mathbf{x}_j)|^2 \frac{Q_\nu}{r_j} = -\frac{Q_\nu^2}{n^2} \tag{23}$$

Therefore, using first the Slater–Condon rules, then (12) and (23),

$$\begin{aligned}\int d\tau\ \mathsf{V}_0(\mathbf{x})|\Phi_\nu(\mathbf{x})|^2 &= -\sum_{\mu\in\nu}\int d\tau_j\ |\chi_\mu(\mathbf{x}_j)|^2\frac{Z}{r_j} \\ &= -\frac{1}{\beta_\nu}\sum_{\mu\in\nu}\int d\tau_j\ |\chi_\mu(\mathbf{x}_j)|^2\frac{Q_\nu}{r_j} \\ &= -\frac{Q_\nu^2}{\beta_\nu}\sum_{\mu\in\nu}\frac{1}{n^2} = \frac{2E_\kappa}{\beta_\nu}\end{aligned} \tag{24}$$

and thus we finally obtain the potential-weighted orthonormality relation:

$$\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_\nu(\mathbf{x}) = \delta_{\nu'\nu}\frac{2E_\kappa}{\beta_\nu} = -\delta_{\nu'\nu}\frac{p_\kappa^2}{\beta_\nu} \tag{25}$$

4 Generalized Sturmian secular equations

We now introduce the definitions

$$T^0_{\nu'\nu} \equiv -\frac{1}{p_\kappa}\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_\nu(\mathbf{x}) \tag{26}$$

and

$$T'_{\nu'\nu} \equiv -\frac{1}{p_\kappa}\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}'(\mathbf{x})\Phi_\nu(\mathbf{x}) \tag{27}$$

From the potential-weighted orthonormality relations (25) and the definition of β_ν (12) it follows that

$$T^0_{\nu'\nu} = \delta_{\nu'\nu}Z\mathcal{R}_\nu \tag{28}$$

Thus the matrix $T^0_{\nu'\nu}$ is diagonal and independent of p_κ. It can be shown ([22], Appendix A) that $T'_{\nu'\nu}$ is also independent of p_κ, although it is not diagonal. We shall call $T'_{\nu'\nu}$ the "energy-independent interelectron repulsion matrix". To obtain the generalized Sturmian secular equations, we begin by substituting the superposition (6) into the Schrödinger equation (3). This yields:

$$\sum_\nu\left[-\frac{1}{2}\Delta + \mathsf{V}(\mathbf{x}) - E_\kappa\right]\Phi_\nu(\mathbf{x})B_{\nu\kappa} = 0 \tag{29}$$

Next, we notice that since all of the isoenergetic Goscinskian configurations in the basis set obey (7), Eq. 29 can be rewritten as

$$\sum_\nu\left[\mathsf{V}(\mathbf{x}) - \beta_\nu\mathsf{V}_0(\mathbf{x})\right]\Phi_\nu(\mathbf{x})B_{\nu\kappa} = 0 \tag{30}$$

We then multiply by a conjugate function from our basis set and integrate over all space and spin coordinates:

$$\sum_\nu\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\left[\mathsf{V}(\mathbf{x}) - \beta_\nu\mathsf{V}_0(\mathbf{x})\right]\Phi_\nu(\mathbf{x})B_{\nu\kappa} = 0 \tag{31}$$

Making use of Eqs. 26–28, we obtain

$$\sum_\nu\left[-p_\kappa\delta_{\nu'\nu}Z\mathcal{R}_\nu - p_\kappa T'_{\nu'\nu} + \beta_\nu p_\kappa\delta_{\nu'\nu}Z\mathcal{R}_\nu\right]B_{\nu\kappa} = 0 \tag{32}$$

Finally, using the relationship

$$\beta_\nu Z\mathcal{R}_\nu = p_\kappa \tag{33}$$

and dividing by p_κ, we obtain the generalized Sturmian secular equations:

$$\sum_\nu \left[\delta_{\nu'\nu} Z\mathcal{R}_\nu + T'_{\nu'\nu} - p_\kappa \delta_{\nu'\nu}\right] B_{\nu\kappa} = 0 \tag{34}$$

Equation 34 differs in several respects from the conventional secular equations used in quantum theory:

1. The kinetic energy term has vanished.
2. The nuclear attraction matrix is diagonal and energy-independent.
3. The interelectron repulsion matrix is energy-independent.
4. The roots are not energies but values of the *scaling parameter*, p_κ, which is proportional to the square roots of the binding energies (Eq. 13).
5. Before the secular equation is solved, only the form of the basis set is known, but not the values of the scaling parameters p_κ. Solution of the secular equations yields a near-optimum basis set appropriate for each state, as well as the states themselves and their corresponding energies.

5 The large-Z approximation: restriction of the basis set to an $\mathcal{R}$-block

If interelectron repulsion is entirely neglected, i.e. when disregarding the second term in Eq. 34, the calculated energies E_κ of Ψ_κ become those of a set of N completely independent electrons moving in the field of the bare nucleus:

$$E_\kappa = -\frac{p_\kappa^2}{2} \longrightarrow -\frac{1}{2}Z^2{\mathcal{R}_\nu}^2 = -\frac{Z^2}{2n_1^2} - \frac{Z^2}{2n_2^2} - \cdots - \frac{Z^2}{2n_N^2} \tag{35}$$

Equation 35 is not the large-Z approximation: in the large-Z approximation, we do not neglect interelectron repulsion, but we restrict the basis set to those Goscinskian configurations that would be degenerate if interelectron repulsion were entirely neglected, i.e., we restrict the basis to a set of configurations all of which correspond to the same value of $\mathcal{R}_\nu$. In that case, the first term in (34) is a multiple of the identity matrix, and the eigenvectors $B_{\nu\kappa}$ are the same as those that would be obtained by diagonalizing the energy-independent interelectron repulsion matrix $T'_{\nu'\nu}$, since the eigenfunctions of any matrix are unchanged by adding a multiple of the unit matrix.

$$\sum_\nu \left[T'_{\nu'\nu} - \lambda_\kappa \delta_{\nu'\nu}\right] B_{\nu\kappa} = 0 \tag{36}$$

The roots are shifted by an amount equal to the constant by which the identity matrix is multiplied:

$$p_\kappa = Z\mathcal{R}_\nu + \lambda_\kappa = Z\mathcal{R}_\nu - |\lambda_\kappa| \tag{37}$$

and the energies become

$$E_\kappa = -\frac{1}{2}(Z\mathcal{R}_\nu - |\lambda_\kappa|)^2 \tag{38}$$

Since the roots λ_κ are always negative, we may use the form $-|\lambda_\kappa|$ in place of λ_κ to make explicit the fact that interelectron repulsion reduces the binding energies, as of course it must. The roots λ_κ are pure numbers that can be calculated once and for all and stored. From these roots, a great deal of information about atomic states can be found with very little effort.

6 Electronic potential at the nucleus

The electronic potential $\varphi(\mathbf{x}_1)$ is related to the electronic density distribution by

$$\varphi(\mathbf{x}_1) = \int d^3x_1' \frac{\rho(\mathbf{x}_1')}{|\mathbf{x}_1 - \mathbf{x}_1'|} \tag{39}$$

If the coordinate system is centered on the nucleus, the electronic potential at the nucleus is then given by

$$\varphi(0) = \int d^3x_1' \frac{\rho(\mathbf{x}_1')}{|\mathbf{x}_1'|} \tag{40}$$

But the electron density corresponding to the state Ψ_κ is defined as

$$\rho(\mathbf{x}_1) = N \int ds_1 \int d\tau_2 \int d\tau_3 \ldots \int d\tau_N \Psi_\kappa^*(\mathbf{x})\Psi_\kappa(\mathbf{x}) \tag{41}$$

where the integral is taken over the spin coordinate of the first electron and over the space and spin coordinates of all the other electrons. The wave function $\Psi_\kappa(\mathbf{x}) = \sum_\nu \Phi_\kappa(\mathbf{x})B_{\nu\kappa}$ is a linear combination of Goscinskian configurations. Thus the density is given by

$$\rho(\mathbf{x}_1) = \sum_{\nu',\nu} \rho_{\nu'\nu}(\mathbf{x}_1) B_{\nu\kappa}^* B_{\nu\kappa} \tag{42}$$

where

$$\begin{aligned}\rho_{\nu'\nu}(\mathbf{x}_1) &= N \int ds_1 \int d\tau_2 \cdots \int d\tau_N \Phi_{\nu'}^*(\mathbf{x})\Phi_\nu(\mathbf{x}) \\ &= \begin{cases} 0 & \text{for} \quad |\nu' \setminus \nu| \geq 2 \\ \chi_{\mu'}^*(\mathbf{x}_1)\chi_\mu(\mathbf{x}_1) & \text{for} \quad \nu' \setminus \nu = \{\mu\} \quad \text{and } \nu \setminus \nu' = \{\mu'\} \\ \sum_{i=1}^N |\chi_{\mu_i}(\mathbf{x}_1)|^2 & \text{for} \quad \nu' = \nu \end{cases}\end{aligned} \tag{43}$$

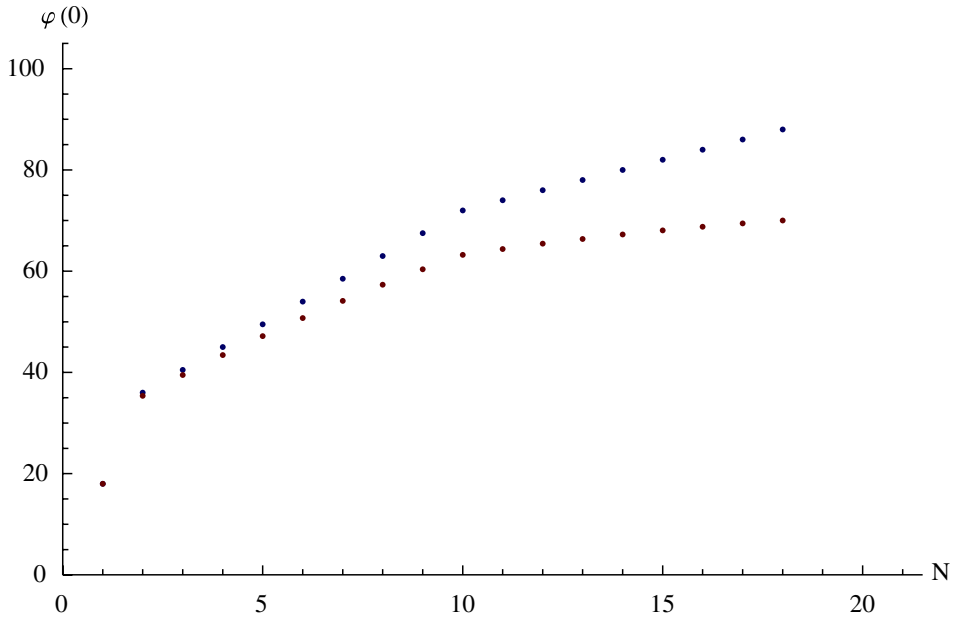


Fig. 1 The electronic potentials at the nuclei of ions in the isonuclear series with nuclear charge $Z = 18$ are shown for $1 \leq N \leq 18$. The upper set of dots were generated neglecting interelectron repulsion, and are seen to be exactly piecewise linear. The lower set of values take interelectron repulsion into account

In Eq. 43 we have made use of the fact that within an $\mathcal{R}$-block, the atomic spin-orbitals are orthonormal.

We calculated the potential at the nucleus (within the framework of the large-Z approximation) by generating the wave functions and from those the densities, using Eqs. 40–43. By this somewhat laborious method, we obtained the values shown in Fig. 1. In this figure, the upper dots correspond to densities generated neglecting interelectron repulsion. In the lower set of dots, interelectron repulsion is included. The fact that the unscreened values were exactly piecewise linear, and that the more exact values that included repulsion were very nearly piecewise linear, was so striking, that we were challenged to explain it.

In finding the explanation we were led to surprisingly simple expressions for the potential at the nucleus of an atom or ion: within the framework of the large-Z approximation we have

$$\begin{aligned}\int d\tau\ \Psi^*_\kappa(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Psi_\kappa(\mathbf{x}) &= \sum_{\nu'}\sum_{\nu} B^*_{\nu'\kappa}B_{\nu\kappa}\int d\tau\ \Phi^*_{\nu'}(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Phi_\nu(\mathbf{x}) \\ &= -\frac{p_\kappa^2}{\beta_\nu}\sum_{\nu}|B_{\nu\kappa}|^2 \end{aligned} \tag{44}$$

In the second step above, we make use of the potential weighted orthonormality relation (25). Further, since $\sum_\nu |B_{\nu\kappa}|^2 = 1$, Eq. 44 reduces to

$$\int d\tau\ \Psi^*_\kappa(\mathbf{x})\mathsf{V}_0(\mathbf{x})\Psi_\kappa(\mathbf{x}) = -\frac{p_\kappa^2}{\beta_\nu} = -p_\kappa Z\mathcal{R}_\nu \tag{45}$$

This result can be used to express the electronic potential at the nucleus in a very simple form. Combining (40) and (41), we obtain

$$\varphi(0) = N \int d\tau \frac{1}{|\mathbf{x}_1|} \Psi_\kappa^*(\mathbf{x}) \Psi_\kappa(\mathbf{x}) \tag{46}$$

From the definition of V_0, (8), and from the fact that each term in the sum in (8) gives the same contribution, we have

$$\varphi(0) = -\frac{1}{Z} \int d\tau \Psi_\kappa^*(\mathbf{x}) \mathsf{V}_0(\mathbf{x}) \Psi_\kappa(\mathbf{x}) \tag{47}$$

Combining Eqs. 45 and 47 we obtain the extremely simple result:

$$\varphi(0) = p_\kappa \mathcal{R}_\nu \tag{48}$$

which can alternatively be written in the form:

$$\varphi(0) = Z\mathcal{R}_\nu{}^2 - |\lambda_\kappa|\mathcal{R}_\nu \tag{49}$$

or in a third form:

$$\varphi(0) = Q_\nu \mathcal{R}_\nu{}^2 \tag{50}$$

since $Q_\nu = Z - |\lambda_\kappa|/\mathcal{R}_\nu$. From Eqs. 48–50 it follows that for an isonuclear series, the electronic potential at the nucleus depends on N in an approximately piecewise linear way. For example, let us consider the isonuclear series where $Z = 18$. Keeping the nuclear charge Z constant at this value, we begin to add electrons. For the ground state we have:

$$\mathcal{R}_\nu{}^2 \equiv \frac{1}{n_1^2} + \frac{1}{n_2^2} + \cdots + \frac{1}{n_N^2} = \begin{cases} \frac{N}{1} & N \le 2 \\ \frac{2}{1} + \frac{N-2}{4} & 2 \le N \le 10 \\ \frac{2}{1} + \frac{8}{4} + \frac{N-10}{9} & 10 \le N \le 18 \end{cases} \tag{51}$$

Tables 1–3 show the roots $|\lambda_\kappa|$ of the energy-independent interelectron repulsion matrix $T'_{\nu'\nu}$. For each value of N, the numerically smallest of these roots corresponds to the ground state. The term $Z\mathcal{R}_\nu{}^2$ is dominant in Eq. 49. When the second term is entirely neglected, i.e., when the effects of interelectron repulsion are neglected, then the dependence of $\varphi(0)$ on N is exactly piecewise linear. However, because of the presence of the second term, the N-dependence is only approximately piecewise linear.

Table 1 Roots of the ground state $\mathcal{R}$-block of the interelectron repulsion matrix for the He-like, Li-like, Be-like, B-like and C-like isoelectronic series

He-like $\lvert\lambda_\kappa\rvert$	Term	Li-like $\lvert\lambda_\kappa\rvert$	Term	Be-like $\lvert\lambda_\kappa\rvert$	Term	B-like $\lvert\lambda_\kappa\rvert$	Term	C-like $\lvert\lambda_\kappa\rvert$	Term
0.441942	^{1}S	0.681870	^{2}S	0.986172	^{1}S	1.40355	^{2}P	1.88151	^{3}P
		0.729017	^{2}P	1.02720	^{3}P	1.44095	^{4}P	1.89369	^{1}D
				1.06426	^{1}P	1.47134	^{2}D	1.90681	^{1}S
				1.09169	^{3}P	1.49042	^{2}S	1.91623	^{5}S
				1.10503	^{1}D	1.49395	^{2}P	1.995141	^{3}D
				1.13246	^{1}S	1.52129	^{4}S	1.96359	^{3}P
						1.54037	^{2}D	1.98389	^{3}S
						1.55726	^{2}P	1.98524	^{1}D
								1.99742	^{1}P
								2.04342	^{3}P
								2.05560	^{1}D
								2.07900	^{1}S

Table 2 Roots of the ground state $\mathcal{R}$-block of the interelectron repulsion matrix $T'_{\nu'\nu}$ for the N-like, O-like, F-like and Ne-like isoelectronic series

N-like $\lvert\lambda_\kappa\rvert$	Term	O-like $\lvert\lambda_\kappa\rvert$	Term	F-like $\lvert\lambda_\kappa\rvert$	Term	Ne-like $\lvert\lambda_\kappa\rvert$	Term
2.41491	^{4}S	3.02641	^{3}P	3.68415	^{2}P	4.38541	^{1}S
2.43246	^{2}D	3.03769	^{1}D	3.78926	^{2}S		
2.44111	^{2}P	3.05065	^{1}S				
2.49314	^{4}P	3.11850	^{3}P				
2.52109	^{2}D	3.14982	^{1}P				
2.53864	^{2}S	3.24065	^{1}S				
2.54189	^{2}P						
2.61775	^{2}P						

In Fig. 2, we see that the values for the electronic potential at the nucleus, laboriously calculated by generating wave functions and densities, are exactly duplicated by the simple expressions in Eqs. 48–50, as well as (51) when interelectron repulsion is neglected.

We should remember that Eqs. 48–50 were derived within the framework of the large-Z approximation. Thus the piecewise-linear dependence of $\varphi(0)$ on N is only an approximate one for two reasons—firstly because the term $-|\lambda_\kappa|\mathcal{R}_\nu$ is slightly nonlinear, and secondly because of the inaccuracies inherent in the large-Z approximation.

Table 3 Numerically smallest roots of the $\mathcal{R}$-blocks and $\mathcal{R}_\nu{}'$-blocks of the interelectron repulsion matrix $T'_{\nu'\nu}$, where $2 \leq N \leq 14$, and where $N' = N - 1$ is the number of electrons remaining after core ionization. Primed quantities refer to the core-ionized states

N	$\mathcal{R}_\nu$	$\lvert\lambda_\kappa\rvert$	Term	N'	$\mathcal{R}_\nu{}'$	$\lvert\lambda'_\kappa\rvert$	Term
2	$\sqrt{8}/2$	0.441942	^{1}S	1	$\sqrt{4}/2$	0.000000	^{2}S
3	$\sqrt{9}/2$	0.681870	^{2}S	2	$\sqrt{5}/2$	0.168089	^{3}S
4	$\sqrt{10}/2$	0.986172	^{1}S	3	$\sqrt{6}/2$	0.433936	^{2}S
5	$\sqrt{11}/2$	1.40355	^{2}P	4	$\sqrt{7}/2$	0.800757	^{3}P
6	$\sqrt{12}/2$	1.88151	^{3}P	5	$\sqrt{8}/2$	1.23703	^{4}P
7	$\sqrt{13}/2$	2.41491	^{4}S	6	$\sqrt{9}/2$	1.73489	^{5}S
8	$\sqrt{14}/2$	3.02641	^{3}P	7	$\sqrt{10}/2$	2.33058	^{4}P
9	$\sqrt{15}/2$	3.68415	^{2}P	8	$\sqrt{11}/2$	2.97391	^{3}P
10	$\sqrt{16}/2$	4.38541	^{1}S	9	$\sqrt{12}/2$	3.66181	^{2}S
11	$\sqrt{37}/3$	4.75240	^{2}S	10	$\sqrt{28}/3$	4.02987	^{3}S
12	$\sqrt{38}/3$	5.1410	^{1}S	11	$\sqrt{29}/3$	4.4243	^{2}S
13	$\sqrt{39}/3$	5.5860	^{2}P	12	$\sqrt{30}/3$	4.8733	^{3}P
14	$\sqrt{40}/3$	6.0512	^{3}P	13	$\sqrt{31}/3$	5.3434	^{4}P

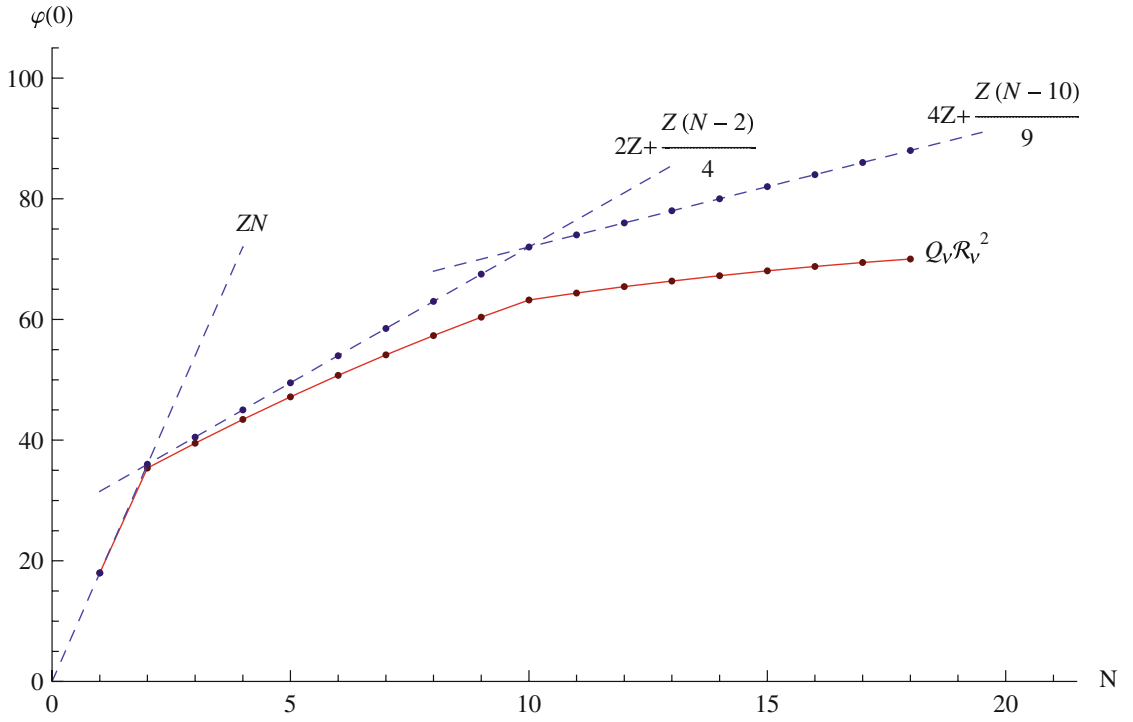


Fig. 2 When interelectron repulsion is entirely neglected, the electronic potential at the nucleus is given by $Z\mathcal{R}_\nu{}^2$, which is exactly piecewise linear in N. The effect of interelectron repulsion is to decrease $\varphi(0)$ and to make the dependence only approximately piecewise linear. The figure shows $\varphi(0)$ neglecting interelectron repulsion (upper values) and including it (lower values). The dots are calculated from the electronic densities of the ground state wave functions, whereas the lines are the closed form expressions found in Eqs. 49 and 51

7 Core ionization energies

The large-Z approximation can be used to calculate core-ionization energies, i.e., the energies required to remove an electron from the inner shell of an atom. From (38) we can see that this energy will be given by

$$\Delta E = \frac{1}{2}\left[(Z\mathcal{R}_\nu - |\lambda_\kappa|)^2 - (Z\mathcal{R}_\nu{}' - |\lambda'_\kappa|)^2\right] \tag{52}$$

where the unprimed quantities refer to the original ground state, while the primed quantities refer to the core-ionized states (Table 2). Since

$$\mathcal{R}_\nu{}^2 - \mathcal{R}_\nu{}'^2 = 1 \tag{53}$$

Equation 52 can be written in the form

$$\Delta E - \frac{Z^2}{2} = Z\left[\mathcal{R}_\nu{}'|\lambda'_\kappa| - \mathcal{R}_\nu|\lambda_\kappa|\right] + \frac{|\lambda_\kappa|^2 - |\lambda'_\kappa|^2}{2} \tag{54}$$

Thus we can see that within the framework of the large-Z approximation, the quantity $\Delta E - Z^2/2$ is linear in Z for an isoelectronic series (Table 3). This quantity represents the contribution of interelectron repulsion to the core ionization energy, since if interelectron repulsion is completely neglected, the core ionization energy is given by $\Delta E = Z^2/2$. Core ionization energies calculated from Eqs. 52–54 are shown in Figs. 3–5.

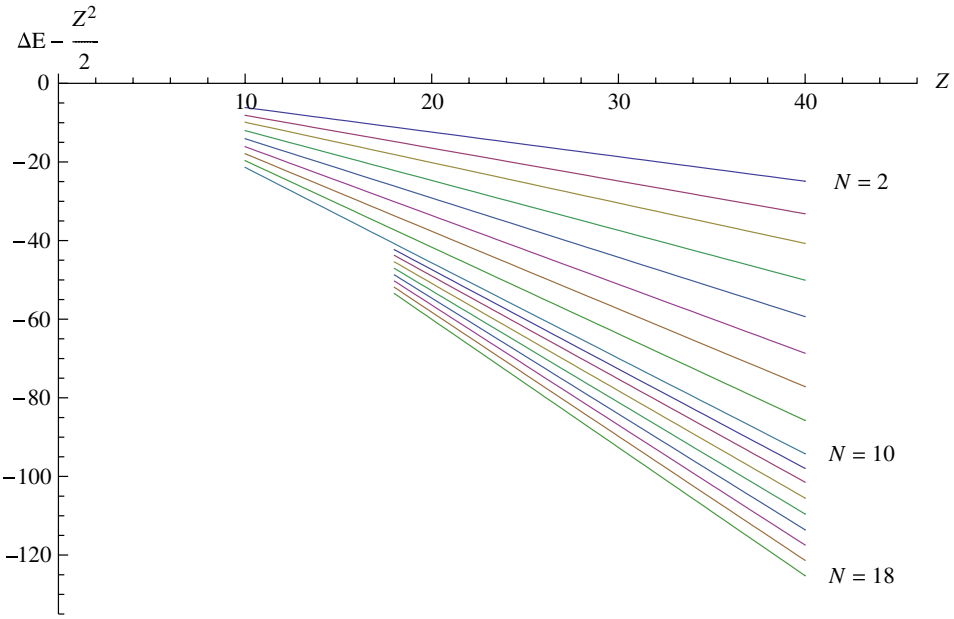


Fig. 3 For isoelectronic series, Eq. 54 indicates that within the large-Z approximation, the quantity $\Delta E - Z^2/2$ is exactly linear in Z, as is illustrated above

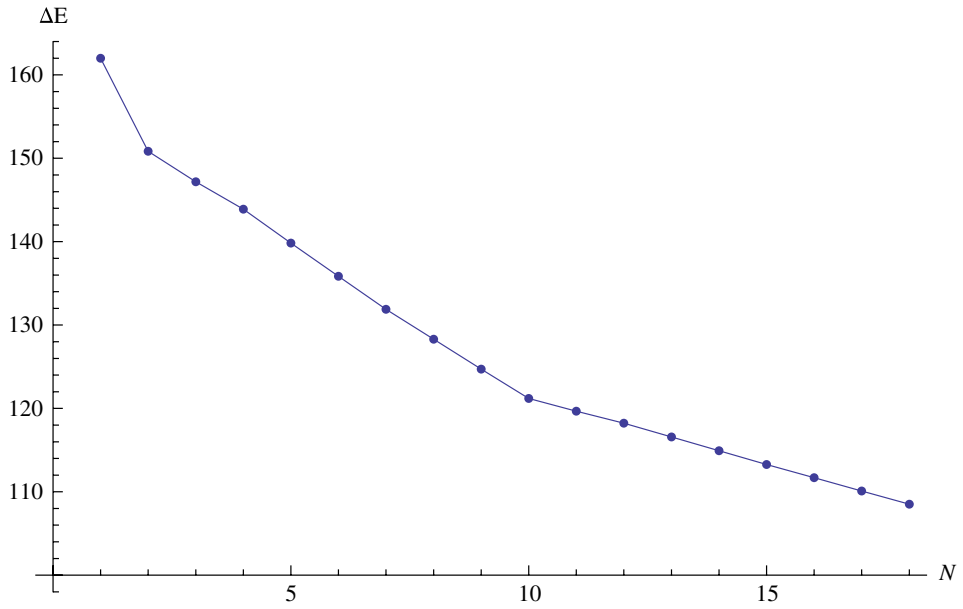


Fig. 4 For isonuclear series, the dependence on N is approximately piecewise linear. Whenever a new shell starts to fill, the slope of the line changes. The dots in the figure were calculated using Eq. 54, where it is not obvious that the dependence ought to be approximately piecewise linear. However, Eqs. 49 and 51 can give us some insight into the approximately piecewise linear relationship

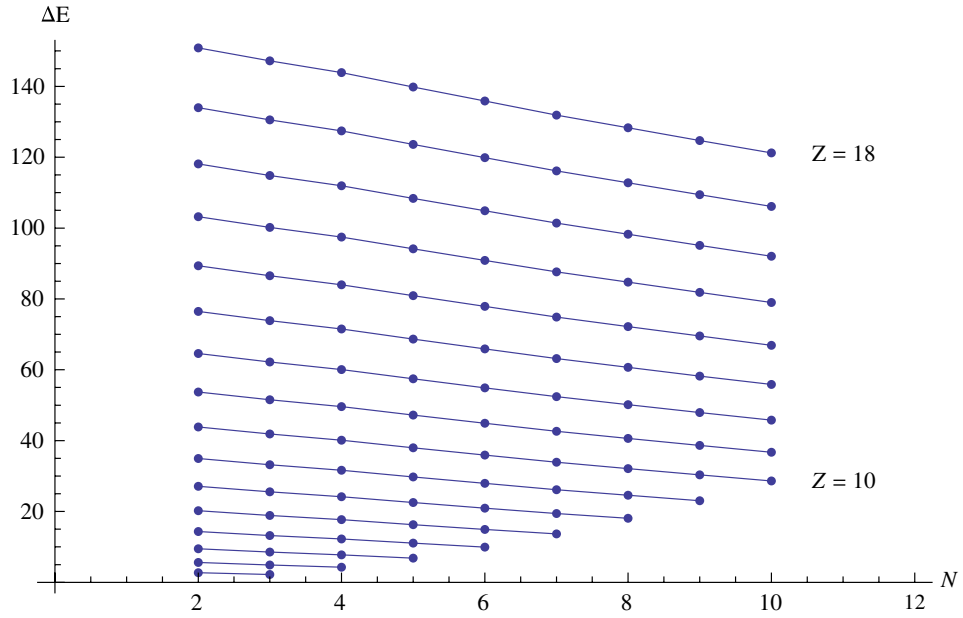


Fig. 5 This figure shows the dependence of the core-ionization energy on both N and Z for the filling of the $n = 2$ shell

8 Validity of the large-Z approximation

In Fig. 6, the large-Z approximation $E_\kappa = -\frac{1}{2}(Z\mathcal{R}_\nu - |\lambda_\kappa|)^2$ for the lowest triplet states of the helium-like isoelectronic series is plotted against spectroscopically determined energies. In order to better see the details, we plot E_κ/Z^2 in Fig. 7. Figure 8 shows E_κ/Z^2 for the ground state of the six-electron isoelectronic series.

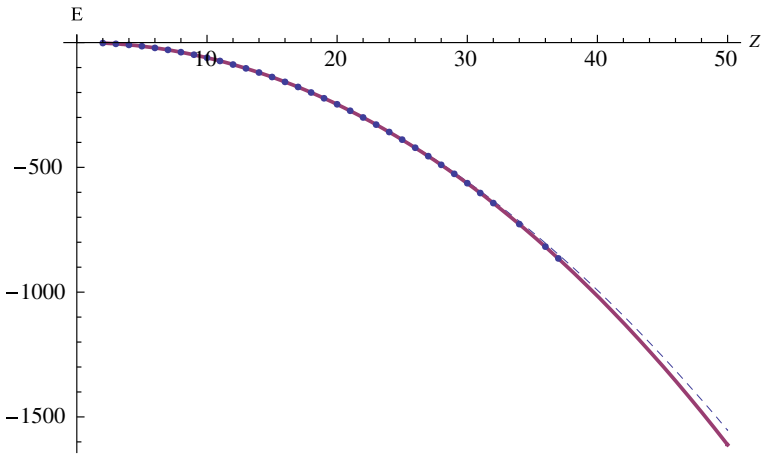


Fig. 6 Energies for the lowest ^{3}S state of the helium-like isoelectronic series calculated in the large-Z approximation, which here limits the basis to a single configuration. The lower (solid) line is corrected for relativistic effects as discussed in the text; the dots indicate experimental values from the NIST tables

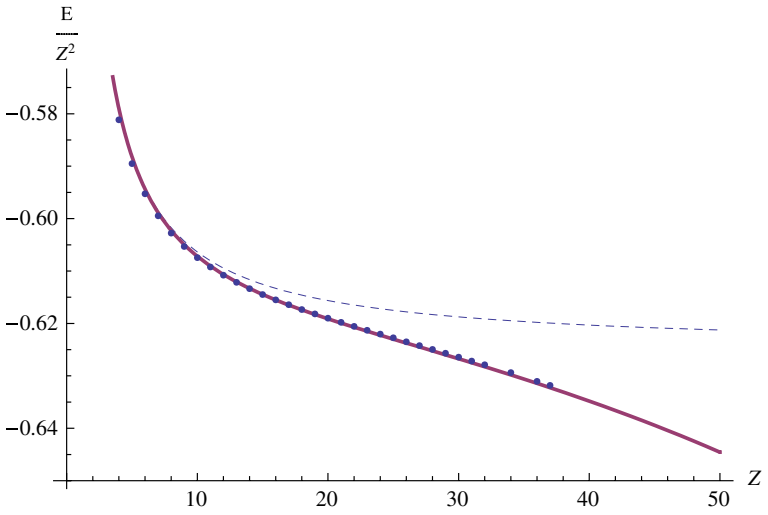


Fig. 7 Here the ^{3}S energies shown in Fig. 6 are divided by Z^2. The lower line is corrected for relativistic effects. The dots are experimental values

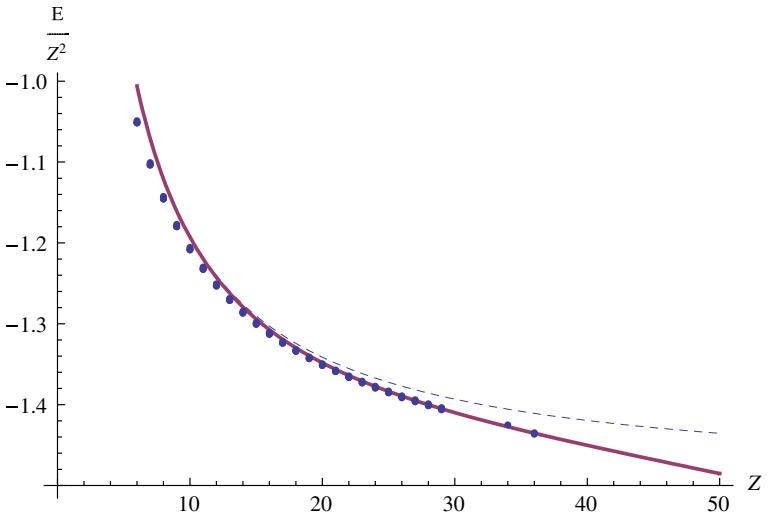


Fig. 8 The ground state of the carbon-like isoelectronic series

As the nuclear charge Z increases, the energies and wave functions calculated with the large-Z approximation approach the exact solutions to the non-relativistic Schrödinger equation. However, relativistic effects begin to be pronounced at around $Z = 10$, and become progressively more so as Z increases. Therefore the calculated values first approach the experimental ones, but begin to differ as relativity becomes more and more important.

It is possible to make a rough correction for the relativistic effect on the energies by multiplying them by an easily calculated factor $f_\kappa(Z)$, so that E_κ becomes

$$-\frac{1}{2} f_\kappa(Z)(Z\mathcal{R}_\nu + \lambda_\kappa)^2 \tag{55}$$

The correction factor $f_\kappa(Z)$ is the ratio between the relativistic and non-relativistic energies of a configuration, assuming interelectron repulsion to be completely neglected such that the energy is equal to that of N independent electrons moving in the field of the nucleus.

In the nonrelativistic case, the energy of a hydrogenlike spin-orbital is given by $-\frac{Z^2}{2n^2}$, and thus the total energy of an N-electron configuration is $-\frac{1}{2}Z^2{\mathcal{R}_\nu}^2$.

In the relativistic case, the energy of a single electron in a hydrogenlike orbital can be found, for example, in [1], or in [22], Eqs. 7.35 through 7.40.

We wish to compare these two energies E_{rel} and E_{nonrel} for a multiconfigurational state

$$\Psi_\kappa = \sum_\nu \Phi_\nu B_{\nu\kappa} \tag{56}$$

the ratio being

$$f_\kappa(Z) = \frac{E_{\text{rel}}}{E_{\text{nonrel}}} = \frac{\sum_\nu B_{\nu\kappa}^2 \langle\Phi_\nu|\mathsf{H}_0|\Phi_\nu\rangle_{\text{rel}}}{\sum_\nu B_{\nu\kappa}^2 \langle\Phi_\nu|\mathsf{H}_0|\Phi_\nu\rangle_{\text{nonrel}}} = \frac{\sum_\nu B_{\nu\kappa}^2 \langle\Phi_\nu|\mathsf{H}_0|\Phi_\nu\rangle_{\text{rel}}}{-\frac{1}{2}Z^2 \sum_\nu B_{\nu\kappa}^2 {\mathcal{R}_\nu}^2} \tag{57}$$

Here, H_0 is a sum of one-electron Hamiltonian operators corresponding to single electrons moving in the field of the bare nucleus, i.e. interelectron repulsion is completely neglected.

In the figures, the lines are calculated in the large-Z approximation. The upper (dashed) line is not corrected for relativistic effects, while the lower (solid) line is corrected. The dots are experimental values of the energies taken from the NIST Atomic Spectra Database [24]. It can be seen from Figs. 5–7 that agreement between the energies calculated from the large-Z approximation and experimental energies become progressively better as Z increases, provided that the rough relativistic correction is made.

We note that the large-Z approximation, despite it's great simplicity, well approximates non-relativistic energies; even for modest values of nuclear charge, the error of the large-Z approximation is smaller than the error due to neglecting relativity. Further, relativistic effects may be accounted for by means of an easily calculated factor, yielding energies that correspond well with experiment.

The second example presented here (the ground state of the carbonlike isoelectronic series) is a case not easily approximated using a small number of Goscinskian basis functions, because interelectron repulsion plays a large role. Nevertheless, it can be seen that even in this somewhat difficult case, the large-Z approximation gives very reasonable results. The large-Z approximation is not only extremely simple, but it is characterized by a small number of parameters–the roots of the interelectron repulsion matrix. These roots are dimensionless and independent of energy and of nuclear charge. They can be calculated once and for all, and they contain information concerning many states of the entire isoelectronic series. Once the roots are obtained, calculating approximate atomic energies, and a number of other properties, become tasks that can be performed on the back of an envelope.

9 Discussion

It can be seen that the core ionization energies predicted from the large-Z approximation depend on N in an approximately piecewise-linear way for isonuclear series, while $\Delta E - Z^2/2$ depends linearly on Z for isoelectronic series. This result is made plausible by Eqs. 48–51, in which the electronic potential at the nucleus is shown to depend approximately piecewise-linearly on N for isonuclear series and linearly on Z for isoelectronic series. It can also be seen that the generalized Sturmian method using Goscinskian configurations provides us with a powerful and convenient method for understanding the properties of atoms and atomic ions. In the large-Z approximation, the method leads to extremely simple closed-form expressions, not only for energies, but also for the behavior of the electronic potential near to the nucleus.

The programs used for the calculations in this paper can be downloaded from: http://sturmian.kvante.org/papers/linear.

References

1. A.I. Akhiezer, V.B. Berestetskii, *Quantum Electrodynamics* (Interscience, New York, 1965)
2. V. Aquilanti, S. Cavalli, D. De Fazio, G. Grossi, in *Hyperangular Momentum: Applications to Atomic and Molecular Science*, ed. by C.A. Tsipis, V.S. Popov, D.R. Herschbach, J.S. Avery. New Methods in Quantum Theory (Kluwer, Dordrecht, 1996)
3. V. Aquilanti, S. Cavalli, C. Coletti, G. Grossi, Alternative Sturmian bases and momentum space orbitals; an application to the hydrogen molecular ion. Chem. Phys. **209**, 405 (1996)
4. J. Avery, *Hyperspherical Harmonics; Applications in Quantum Theory* (Kluwer Academic Publishers, Dordrecht, 1989)
5. J. Avery, D.R. Herschbach, Hyperspherical Sturmian basis functions. Int. J. Quantum Chem. **41**, 673 (1992)
6. J. Avery, in *Hyperspherical Harmonics; Some Properties and Applications*, ed. by E.S. Kryachko, J.L. Calais. Conceptual Trends in Quantum Chemistry (Kluwer, Dordrecht, 1994)
7. J. Avery, Many-particle Sturmians. J. Math. Chem. **21**, 285 (1997)
8. J. Avery, A formula for angular and hyperangular integration. J. Math. Chem. **24**, 169 (1998)
9. J. Avery, *Hyperspherical Harmonics and Generalized Sturmians* (Kluwer Academic Publishers, Dordrecht, Netherlands, 2000), p. 196
10. J. Avery, S. Sauer, in *Many-electron Sturmians Applied to Molecules*, ed. by A. Hernández-Laguna, J. Maruani, R. McWeeney, S. Wilson. Quantum Systems in Chemistry and Physics, Vol. 1 (Kluwer Academic Publishers, 2000)

11. J. Avery, in *Sturmians*, ed. by S. Wilson. Handbook of Molecular Physics and Quantum Chemistry (Wiley, Chichester, 2003)
12. J. Avery, Many-center Coulomb Sturmians and Shibuya-Wulfman integrals. Int. J. Quantum Chem. (2003)
13. J. Avery, J. Avery, O. Goscinski, Natural orbitals from generalized Sturmian calculations. Adv. Quantum Chem. **43**, 207–216 (2003)
14. J. Avery, J. Avery, The generalized Sturmian method for calculating spectra of atoms and ions. J. Math. Chem. **33**, 145–162 (2003)
15. J. Avery, J. Avery, V. Aquilanti, A. Caligiana, Atomic densities, polarizabilities and natural orbitals derived from generalized Sturmian calculations. Adv. Quantum Chem. **47** 156–173 (2004)
16. J. Avery, J. Avery, Generalized Sturmian solutions for many-particle Schrödinger equations. J. Phys. Chem. A **41**, 8848 (2004)
17. J. Avery, J. Avery, Autoionizing states of atoms calculated using generalized Sturmians. Adv. Quantum Chem. **49**, 103–118 (2005)
18. A. Caligiana, Sturmian Orbitals in Quantum Chemistry, Ph.D. thesis, University of Perugia, Italy, October 2003
19. O. Goscinski, Adv. Quantum Chem. **41**, 51–85 (2003)
20. T. Koga, T. Matsuhashi, Sum rules for nuclear attraction integrals over hydrogenic orbitals. J. Chem. Phys. **87**(8), 4696–4699 (1987)
21. T. Koga, T. Matsuhashi, One-electron diatomics in momentum space. V. Nonvariational LCAO approach. J. Chem. Phys. **89**, 983 (1988)
22. J. Avery, J. Avery, *Generalized Sturmians and Atomic Spectra* (World Scientific, 2007)
23. J. Avery, J. Avery, The generalized Sturmian library, http://sturmian.kvante.org (2006)
24. National Institute of Standards and Technology (NIST), NIST Atomic Spectra Database, http://physics.nist.gov/asd.

Publications: Other

Static complexity analysis of higher order programs

Published in FOPARA'09 Proceedings of the First international conference on Foundational and practical aspects of resource analysis, Springer-Verlag: Avery et al. [2010].

Static Complexity Analysis of Higher Order Programs

James Avery[1], Lars Kristiansen[2], and Jean-Yves Moyen[3⋆]

[1] DIKU, University of Copenhagen
[2] Department of Mathematics, University of Oslo
[3] LIPN - UMR 7030 – CNRS - Université Paris 13 – F-93430 Villetaneuse, France

Abstract. The overall goal of the research presented in this paper is to find automatic methods for static complexity analysis of higher order programs.

1 Introduction and Related Work

In the first part of the paper we consider a first order imperative language. Our method for dealing with this language is reminiscent of the methods of Ben-Amram, Jones and Kristiansen [JK09,KJ05,BAJK08] or of Niggl and Wünderlich [NW06]. The method is also, although the link is not obvious, related to quasi-interpretation and sup-interpretation; see e.g. Bonfante, Marion and Moyen [BMM07] and Marion and Péchoux [MP09].

In the second part of the paper, we lift the method developed in the first part to allow analysis of higher order programs written in a language that is an imperative version of Gödel's T. There have not been much research on automatic complexity analysis of higher order programs. One exception is Benzinger [B04], however, the semi-automatic approach taken in [B04] seems to be very different from our approach. We believe the work presented in this paper suggests how a number of methods, including the powerful method in [BAJK08], can be generalised to deal with higher order programs.

The paper presents work in progress. The main result is still conjecture, and there is much room for simplification, but several examples are given that illustrate why and how the methods should work.

2 Vectors and Matrices over Semirings

Recall that a *semiring* is a set $\mathcal{S}$ together with two internal operations $+$ and $\cdot$, called *addition* and *multiplication*, such that: $+$ is associative, commutative and has a neutral element 0; $\cdot$ is associative and has a neutral element 1; 0 annihilates $\cdot$; $\cdot$ is distributive over $+$.

⋆ This work was partially supported by a grant from Agence Nationale de la Recherche, ref. ANR-08-BLANC-0211-01 (COMPLICE project)

We consider vectors over a semiring $\mathcal{S}$. To make working with sparse representations easier, we associate an index set $\mathcal{I}$ and define a vector as a mapping $V : \mathcal{I} \to \mathcal{S}$. This representation is equivalent to the standard one, in which vectors are represented by $\mathcal{S}$-tuples denoting the coefficients in a linear expansion in a basis. $\mathcal{I}$ might be infinite, but in our analysis, any vector will have only finitely many nonzero entries.

If $i \in \mathcal{I}$ is an index and V is a vector, $V[i]$ denotes the i^{th} component of the vector (that is, V_i in usual notation). Addition and scalar multiplication are canonically lifted from scalars to vectors point-wise: $(V + W)[i] = V[i] + W[i]$ and $(a \cdot V)[i] = a \cdot (V[i])$.

We use the sparse representation $\left(\begin{smallmatrix} x_1 & \cdots & x_n \\ a_1 & \cdots & a_n \end{smallmatrix}\right)$ to denote the vector V, where $V[x_i] = a_i$ for $\vec{a} \in \mathcal{S}$ and $\vec{x} \in \mathcal{I}$, and where $V[y] = 0$ for any $y \in \mathcal{I} \setminus \{x_1, \ldots, x_n\}$. Notice that $()$ then is the zero vector and that

$$\begin{pmatrix} x_1 & \cdots & x_i \\ a_1 & & a_i \end{pmatrix} \oplus \begin{pmatrix} x_{i+1} & \cdots & x_n \\ a_{i+1} & & a_n \end{pmatrix} = \begin{pmatrix} x_1 & \cdots & x_n \\ a_1 & & a_n \end{pmatrix}$$

when $\{x_1, \ldots, x_i\}$ and $\{x_{i+1}, \ldots, x_n\}$ are disjoint. Let V be a vector. We define

$$\begin{pmatrix} x_1 & \cdots & V & \cdots & x_n \\ a_1 & & a_i & & a_n \end{pmatrix} = a_i V \oplus \begin{pmatrix} x_1 & \cdots & x_{i-1} & x_{i+1} & \cdots & x_n \\ a_1 & & a_{i-1} & a_{i+1} & & a_n \end{pmatrix} \qquad (*)$$

This notation will turn out to be convenient. We will use $e\,[x \setminus t]$ to denote the expression e where the free occurrences of x have been replaced by t. By (*),

$$\begin{pmatrix} x_1 & \cdots & x_i & \cdots & x_n \\ a_1 & & a_i & & a_n \end{pmatrix} [x_i \setminus V] = a_i V \oplus \begin{pmatrix} x_1 & \cdots & x_{i-1} & x_{i+1} & \cdots & x_n \\ a_1 & & a_{i-1} & a_{i+1} & & a_n \end{pmatrix} \qquad (**)$$

We will also see that the equality (**) provides a convenient way of defining the product of higher order matrices by certain substitutions.

A matrix over scalars $\mathcal{S}$ and index set $\mathcal{I}$ is seen as a mapping from the indexes to vectors (over $\mathcal{S}$ and $\mathcal{I}$). These vectors are the usual column vectors of algebra textbooks. We use $M, A, B, C, \ldots$ to denote matrices, and M_j to denote the vector that the matrix M assigns to the index j. Thus, $M_j[i]$ is a scalar.

We have chosen to use the nonstandard notation $M_j[i]$ in place of the standard notation M_{ij}, the reason being that this notation works much better for higher order. The reader should note that M_j denotes the j^{th} column vector of the matrix M, i.e. M_j^T in standard row-major notation.

Sums and *products* of matrices are defined as usual:

$$(A + B)_j[i] = A_j[i] + B_j[i] \qquad (A \cdot B)_j[i] = \sum_{k \in \mathcal{I}} A_k[i] \cdot B_j[k]$$

The matrix M is an *upper bound* of the matrix A, and we write $M \geq A$, iff there exists a matrix B such that $M = A + B$. Thus we have a partial ordering of matrices. The ordering symbols $\geq, \leq, >, <$ have their standard meaning with respect to this ordering, and we will use standard terminology, that is, we say that A is greater than B when $A \geq B$, that A is strictly less than B when $A < B$, et cetera.

The *zero matrix* is denoted by $\mathbf{0}$, and $M = \mathbf{0}$ iff $M_j = ()$ for all $j \in \mathcal{I}$. We have $\mathbf{0}+M = M+\mathbf{0} = M$ for any matrix M. The *identity matrix* is denoted by $\mathbf{1}$, and $M = \mathbf{1}$ iff $M_j[i] = 1$ for $i = j$, $M_j[i] = 0$ for $i \neq j$. We have $\mathbf{1} \cdot M = M \cdot \mathbf{1} = M$ for any matrix M. Furthermore, let $M^0 = \mathbf{1}$ and $M^{n+1} = M \cdot M^n$.

We define the *closure* of the matrix M, written M^* by the infinite sum

$$M^* \;=\; \mathbf{1} \;+\; M \;+\; M^2 \;+\; M^3 \;+\; \ldots$$

Let $\mathcal{M}_n$ denote the set of $n \times n$ matrices over a closed semiring. Then the algebraic structure $(\mathcal{M}_n, +, \cdot, \mathbf{0}, \mathbf{1})$ is a semiring. The closure M^* may not exist for every matrix $M \in \mathcal{M}_n$. However, if M^* exists, the identity $M^* = \mathbf{1}+(M \cdot M^*)$ holds.

3 Analysis of a First-Order Programming Language

3.1 The programming language

Syntax We consider deterministic imperative programs that manipulate natural numbers held in variables. Each variable stores a single natural number. Our language is an extension of the well-know Loop language studied in Meyer & Ritchie [MR76] and in several other places. In this language, a function is computable if and only if it is primitive recursive.

Definition. *Expressions* and *sequences* are defined by the following grammar:

$$\begin{array}{lll}
\text{(Variables)} & & ::= \mathtt{X} \\
\text{(Constants)} & & ::= k_n \text{ for each } n \in \mathbb{N} \\
\text{(Operators)} & & ::= \mathtt{op} \\
\text{(Expressions)} & \ni e & ::= \mathtt{op}(e_1, e_2) \mid \mathtt{X} \mid k_n \\
\text{(Sequences)} & \ni s & ::= \varepsilon \mid s_1\,\mathtt{;}\,s_2 \mid \mathtt{X:=}\,e \mid \mathtt{loop\ X}\,\{\,s\,\}
\end{array}$$

Sequences will also be called *programs*.

Semantics A program is executed as expected from its syntax, so we omit a detailed formalisation. The semantics at first order is the restriction of the high-order semantics depicted in Figure 1. At any execution step, each variable $\mathtt{X}_i$ holds a natural number x_i, and the expressions are evaluated straightforwardly without side effects; there is an unspecified set of operators that are all computable in constant time. Typical operators include `add`, `mul`, `sub` and `max`, whose semantics are respectively addition, multiplication, modified subtraction[4] and maximum. The constant k_n denotes the integer $n \in \mathbb{N}$. The program $\mathtt{loop\ X}\,\{\,s\,\}$ executes the sequence s in its body x times in a row, where x is the value stored in `X` when the loop starts. `X` may not appear in the body of the loop. The sequence $s_1\,\mathtt{;}\,s_2$ executes first the sequence s_1 followed by the sequence s_2. Programs of the form $\mathtt{X:=}\,e$ are ordinary assignment statements, and the command ε does nothing. In all, the semantics of the language is straightforward.

[4] That is, $\mathtt{sub}(\mathtt{X}, \mathtt{Y})$ returns 0 if `Y` is greater than `X`.

Feasible Programs Let s be a program with variables $\{\mathtt{X}_1, \ldots, \mathtt{X}_n\}$. The program execution relation $a_1, \ldots, a_n[s]b_1, \ldots, b_n$ holds iff the variables $\mathtt{X}_1, \ldots, \mathtt{X}_n$ respectively hold the numbers $a_1, \ldots, a_n$ when the execution of s starts and respectively the numbers $b_1, \ldots, b_n$ when the execution terminates. We say that a program s is *feasible* if for any subprogram s' there exist polynomials $p_1, \ldots, p_n$ such that $a_1, \ldots, a_n[s']b_1, \ldots, b_n \;\Rightarrow\; b_i \leq p_i(a_1, \ldots, a_n)$.

3.2 Abstract Interpretation

A Particular Semiring We will analyse programs by interpreting expressions as vectors and sequences as matrices. These vectors and matrices will be over

- the index set $k, \mathtt{X}_0, \mathtt{X}_1, \mathtt{X}_2, \ldots$, that is, the set of program variables extended by the index k,
- and the semiring $(\mathbb{N}, \oplus, \otimes, 0, 1)$ where $\oplus$ is the maximum operator, i.e. $a \oplus b = \max(a, b)$, and $\otimes$ is standard multiplication of natural numbers.

Example In this semiring, the closure M^* of a matrix M might not exists. Let A and B be the following matrices over $(\mathbb{N}, \oplus, \otimes, 0, 1)$ and the index set $\{1, 2, 3\}$:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \qquad\qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then we have

- $A^2 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & 2 & 1 \end{smallmatrix}\right)$, $A^3 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2 & 1 \end{smallmatrix}\right)$, and $A^4 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2 & 1 \end{smallmatrix}\right)$
- $B^2 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 4 & 4 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$, $B^3 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 8 & 8 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$, and $B^3 = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 16 & 16 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$

Hence, we see that A^* exists as

$$A^* \;=\; \mathbf{1} \oplus A \oplus A^2 \oplus A^3 \oplus \ldots \;=\; \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2 & 1 \end{pmatrix}$$

whereas $B^* = \left(\begin{smallmatrix} 1 & 0 & 0 \\ \infty & \infty & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$ is not a matrix over $\mathbb{N}$.

Some Intuition Our goal is to decide whether or not a given program is feasible by analysing the syntax of the program. This feasibility problem is of course undecidable, but we will present a sound computable method in the sense that if the method certifies a program as feasible, the program will indeed be feasible.

Our abstract interpretation $[\![s]\!]$ of a program s will either be undefined or a matrix

$$[\![s]\!] \;=\; [\mathtt{X}_1 \mapsto V_1,\; \ldots, \mathtt{X}_n \mapsto V_n]$$

where $\mathtt{X}_1, \ldots, \mathtt{X}_n$ are the variables occurring in s and $V_1, \ldots, V_n$ are vectors over $\mathbb{N}$ and the index set $\{\mathtt{X}_1, \ldots, \mathtt{X}_n\}$. The program s will be feasible if $[\![s]\!]$ is defined, and the vector V_i will say something about how the output value of $\mathtt{X}_i$ depends on the input values of $\mathtt{X}_1, \ldots, \mathtt{X}_n$. Let $x_1, \ldots, x_n$ be the input values, and let $x'_1, \ldots, x'_n$ be the output values, of respectively $\mathtt{X}_1, \ldots, \mathtt{X}_n$. If $V_i[\mathtt{X}_j] = 0$, then

there exists a polynomial independent of x_j such that $x'_i < p$. If $V_i[\mathtt{X}_j] = 1$, then there exists a polynomial q independent of x_j such that $x'_i \leq x_j + q$. Finally, if $V_i[\mathtt{X}_j] > 1$, then there exists a polynomial p dependent of x_j such that $x'_i \leq p$. The exact value of $V_i[\mathtt{X}_j]$ will say something about the degree of the variable x_j in this polynomial: if $V_i[\mathtt{X}_j] = k$, then the degree of x_j will be less or equal to k. If $[\![s]\!]$ is undefined, s might not be a feasible program.

Let us study an example where s is the program $\mathtt{loop\ U\ \{\ X := add(X, Y);\ Z := X\ \}}$. Our interpretation of s will be

$$[\![s]\!] = \left[\mathtt{X} \mapsto \binom{\mathtt{XYZU}}{1202},\ \mathtt{Y} \mapsto \binom{\mathtt{XYZU}}{0100},\ \mathtt{Z} \mapsto \binom{\mathtt{XYZU}}{1202},\ \mathtt{U} \mapsto \binom{\mathtt{XYZU}}{0001}\right].$$

This interpretation tells something about the program s. The entry $\mathtt{X} \mapsto \binom{\mathtt{XYZU}}{1202}$ ensures that there exists a polynomial $p(y, u)$ such that the output value of $\mathtt{X}$ is bounded by $x + p(y, u)$, where x, y, u are the input values of respectively $\mathtt{X}, \mathtt{Y}, \mathtt{U}$. We cannot read off an exact polynomial, but we can conclude that such polynomial exists. Furthermore, we can conclude that the polynomial may be *dependent* of the input values of $\mathtt{Y}$ and $\mathtt{U}$ and is *independent* of the input values of $\mathtt{X}$ and $\mathtt{Z}$. By inspecting the program s, we can see that $p(y, u)$ might be the polynomial $y \times u$, but we cannot deduce this information from the interpretation $[\![s]\!]$. However, we know from $[\![s]\!]$ that there exists *some* polynomial $p(y, u)$ where the degrees of y and u is less or equal to 2. The entry $\mathtt{Y} \mapsto \binom{\mathtt{XYZU}}{0100}$ tells us that the output value of $\mathtt{Y}$ is bounded by the input value of $\mathtt{Y}$ and independent of the input values of the remaining program variables. The interpretation $[\![s]\!]$ yields similar information about the output values of $\mathtt{Z}$ and $\mathtt{U}$.

The interpretation $[\![s]\!]$ will be undefined for any infeasible program s. E.g., the interpretation of the program $\mathtt{loop\ X\ \{\ Y := mul(Y, Y)\ \}}$ will not be defined as the output value of $\mathtt{Y}$ will not be bounded by a polynomial in the input values of $\mathtt{X}$ and $\mathtt{Y}$. We have

$$\begin{aligned}
[\![\mathtt{Y := mul(Y, Y)}]\!] &= [\mathtt{X} \mapsto \tbinom{\mathtt{XY}}{10}, \mathtt{Y} \mapsto \tbinom{\mathtt{XY}}{02}] \\
[\![\mathtt{Y := mul(Y, Y);\ Y := mul(Y, Y)}]\!] &= [\mathtt{X} \mapsto \tbinom{\mathtt{XY}}{10}, \mathtt{Y} \mapsto \tbinom{\mathtt{XY}}{04}] \\
[\![\mathtt{Y := mul(Y, Y);\ Y := mul(Y, Y);\ Y := mul(Y, Y)}]\!] &= [\mathtt{X} \mapsto \tbinom{\mathtt{XY}}{10}, \mathtt{Y} \mapsto \tbinom{\mathtt{XY}}{08}] \\
&\ \ \vdots
\end{aligned}$$

but $[\![\mathtt{loop\ X\ \{\ Y := mul(Y, Y)\ \}}]\!]$ is undefined.

Loop Correction In addition to the standard operators on vectors and matrices defined in Section 2, we need the operator of *loop correction*: For each $\mathtt{X} \in \mathcal{I}$, we define a unary operation $M^{\downarrow \mathtt{X}}$ on a matrix M, by

$$M_j^{\downarrow \mathtt{X}} = \begin{cases} M_j \oplus \binom{\mathtt{X}}{a} & \text{if } a > 1 \\ M_j & \text{otherwise} \end{cases}$$

where $a = \sum_{i \in \mathcal{I}} M_j[i]$. Note that both the closure and the loop correction operator are monotonous, i.e. $M \leq M^*$ (if M^* exists) and $M \leq M^{\downarrow \mathtt{X}}$ for any matrix M and any $\mathtt{X} \in \mathcal{I}$.

The Interpretation Operator We define the interpretation operator $[\![\cdot]\!]$ mapping expressions to vectors and sequences to matrices.

- Interpretations of expressions:
  - for any program variable $\mathtt{X}$, let $[\![\mathtt{X}]\!] = \binom{\mathtt{X}}{1}$.
  - for any constant $\mathtt{k}_n$, let $[\![\mathtt{k}_n]\!] = \binom{k}{1}$.
  - $[\![\mathtt{op}(e_1, e_2)]\!] = [\![\mathtt{op}]\!]([\![e_1]\!], [\![e_2]\!])$, i.e. the function $[\![\mathtt{op}]\!]$ applied to the arguments $[\![e_1]\!]$ and $[\![e_2]\!]$.
  - $[\![\mathtt{add}]\!] = \lambda x \lambda y \left(\binom{x}{1} \oplus \binom{y}{2}\right) = \lambda x \lambda y \begin{pmatrix} x & y \\ 1 & 2 \end{pmatrix}$.
  - $[\![\mathtt{mul}]\!] = \lambda x \lambda y \left(\binom{x}{2} \oplus \binom{y}{2}\right) = \lambda x \lambda y \begin{pmatrix} x & y \\ 2 & 2 \end{pmatrix}$
  - $[\![\mathtt{sub}]\!] = \lambda x \lambda y \begin{pmatrix} x \\ 1 \end{pmatrix}$
  - $[\![\mathtt{max}]\!] = \lambda x \lambda y \left(\binom{x}{1} \oplus \binom{y}{1}\right) = \lambda x \lambda y \begin{pmatrix} x & y \\ 1 & 1 \end{pmatrix}$
- Interpretations of sequences:
  - $[\![s_1 \,;\, s_2]\!] = [\![s_1]\!] \otimes [\![s_2]\!]$
  - $[\![\mathtt{X} := e]\!] = \mathbf{1}^{\mathtt{X}}_{[\![e]\!]}$ where $\mathbf{1}^{\mathtt{X}}_{[\![e]\!]}$ denotes the identity matrix $\mathbf{1}$ where the vector indexed by $\mathtt{X}$ is replaced by the vector $[\![e]\!]$
  - $[\![\varepsilon]\!] = \mathbf{1}$ (the identity matrix)
  - $[\![\mathtt{loop}\ \mathtt{X}\ \{\ s\ \}]\!] = \begin{cases} ([\![s]\!]^*)^{\downarrow \mathtt{X}} & \text{if } [\![s]\!]^* \text{ exists} \\ \text{undefined} & \text{otherwise} \end{cases}$

This interpretation is based on the following idea: Let s be a program whose variables are a subset of $\{\mathtt{X}_1, \ldots, \mathtt{X}_n\}$. There might exist M such that $[\![s]\!] = M$, and there might not exist such an M. If M exists, then there exists polynomials $p_1, \ldots, p_n$ such that

$$a_1, \ldots, a_n[s]b_1, \ldots, b_n \quad \Rightarrow \quad b_i \leq \max(\vec{u}) + p_i(\vec{v})$$

where a_j is in the list $\vec{u}$ iff $M_{\mathtt{X}_i}[\mathtt{X}_j] = 1$; and where a_j is in the list $\vec{v}$ iff $M_{\mathtt{X}_i}[\mathtt{X}_j] > 1$. Hence, if $M_{\mathtt{X}_i}[\mathtt{X}_j] = 0$, a bound on b_i does not depend on a_j. If M does not exist, then some output value b_i might not be polynomially bounded by the input values $a_1, \ldots, a_n$.

Example We will compute the the interpretation of the program expression $\mathtt{max}(\mathtt{X}_1, \mathtt{X}_2)$. Recall definition (*) from Section 2, that is, for $\vec{x} \in \mathcal{I}$ and $\vec{a} \in \mathcal{S}$ and any vector V, we have

$$\begin{pmatrix} x_1 & \cdots & V & \cdots & x_n \\ a_1 & \cdots & a_i & \cdots & a_n \end{pmatrix} = a_i V \oplus \begin{pmatrix} x_1 & \cdots & x_{i-1} & x_{i+1} & \cdots & x_n \\ a_1 & \cdots & a_{i-1} & a_{i+1} & \cdots & a_n \end{pmatrix}$$

$$\begin{aligned} [\![\mathtt{max}(\mathtt{X}_1, \mathtt{X}_2)]\!] &= \lambda x \lambda y \begin{pmatrix} x & y \\ 1 & 1 \end{pmatrix} \binom{\mathtt{X}_1}{1} \binom{\mathtt{X}_2}{1} = \lambda y \begin{pmatrix} \binom{\mathtt{X}_1}{1} & y \\ 1 & 1 \end{pmatrix} \binom{\mathtt{X}_2}{1} \\ &= \lambda y \left(1 \binom{\mathtt{X}_1}{1} \oplus \binom{y}{1}\right) \binom{\mathtt{X}_2}{1} = \lambda y \begin{pmatrix} \mathtt{X}_1 & y \\ 1 & 1 \end{pmatrix} \binom{\mathtt{X}_2}{1} = \begin{pmatrix} \mathtt{X}_1 & \binom{\mathtt{X}_2}{1} \\ 1 & 1 \end{pmatrix} \\ &= 1 \binom{\mathtt{X}_2}{1} \oplus \binom{\mathtt{X}_1}{1} = \begin{pmatrix} \mathtt{X}_1 & \mathtt{X}_2 \\ 1 & 1 \end{pmatrix} \end{aligned}$$

Example Next we compute the the interpretation of the program expression $\mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)$. We have

$$\begin{aligned}
[\![\mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)]\!] &= \lambda x\lambda y \begin{pmatrix} x & y \\ 2 & 2 \end{pmatrix} ([\![\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2)]\!], [\![\mathtt{X}_3]\!]) \\
&= \lambda x\lambda y \begin{pmatrix} x & 2 \\ 2 & 2 \end{pmatrix} (\lambda x\lambda y \begin{pmatrix} x \\ 1 \end{pmatrix} ([\![\mathtt{X}_1]\!], [\![\mathtt{X}_2]\!]), [\![\mathtt{X}_3]\!]) \\
&= \lambda x\lambda y \begin{pmatrix} x & y \\ 2 & 2 \end{pmatrix} \left(\begin{pmatrix} [\![\mathtt{X}_1]\!] \\ 1 \end{pmatrix}, [\![\mathtt{X}_3]\!] \right) = \lambda x\lambda y \begin{pmatrix} x & y \\ 2 & 2 \end{pmatrix} \left(\begin{pmatrix} \begin{pmatrix} \mathtt{X}_1 \\ 1 \end{pmatrix} \\ 1 \end{pmatrix}, \begin{pmatrix} \mathtt{X}_3 \\ 1 \end{pmatrix} \right) \\
&= \lambda x\lambda y \begin{pmatrix} x & y \\ 2 & 2 \end{pmatrix} \left(\begin{pmatrix} \mathtt{X}_1 \\ 1 \end{pmatrix}, \begin{pmatrix} \mathtt{X}_3 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} \begin{pmatrix} \mathtt{X}_1 \\ 1 \end{pmatrix} & \begin{pmatrix} \mathtt{X}_3 \\ 1 \end{pmatrix} \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 2 & 2 \end{pmatrix}
\end{aligned}$$

The computation shows that the expression $\mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)$ is interpreted as the vector $\begin{pmatrix} \mathtt{X}_1 & \mathtt{X}_2 \\ 2 & 2 \end{pmatrix}$. Hence, there exists a polynomial p such that the value of the expression is bounded by $p(\mathtt{X}_1,\mathtt{X}_3)$. Note that the very value of the expression indeed also depends on $\mathtt{X}_2$, however, there exists a polynomial bound on this value which does not depend on $\mathtt{X}_2$, e.g. the polynomial $\mathtt{X}_1 \times \mathtt{X}_3$.

Example We will now show how to compute the interpretation of the program

$$\mathtt{X}_3 := \mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)\,;\, \mathtt{X}_2 := \mathtt{max}(\mathtt{X}_2,\mathtt{X}_3)\ .$$

The interpretations of the expressions occurring in the program are computed in the previous examples, and we have

$$\begin{aligned}
&[\![\mathtt{X}_3 := \mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)\,;\, \mathtt{X}_2 := \mathtt{max}(\mathtt{X}_2,\mathtt{X}_3)]\!] \;= \\
&\qquad [\![\mathtt{X}_3 := \mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)]\!] \otimes [\![\mathtt{X}_2 := \mathtt{max}(\mathtt{X}_2,\mathtt{X}_3)]\!] \;= \\
&\qquad\qquad \mathbf{1}^{\mathtt{X}_3}_{[\![\mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)]\!]} \otimes \mathbf{1}^{\mathtt{X}_2}_{[\![\mathtt{max}(\mathtt{X}_2,\mathtt{X}_3)]\!]} \;=\; \mathbf{1}^{\mathtt{X}_3}_{\left(\begin{smallmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 2 & 2 \end{smallmatrix}\right)} \otimes \mathbf{1}^{\mathtt{X}_2}_{\left(\begin{smallmatrix} \mathtt{X}_2 & \mathtt{X}_3 \\ 1 & 1 \end{smallmatrix}\right)}
\end{aligned}$$

The computation shows that the interpretation of the program is the product of the the two matrices $\mathbf{1}^{\mathtt{X}_3}_{\left(\begin{smallmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 2 & 2 \end{smallmatrix}\right)}$ and $\mathbf{1}^{\mathtt{X}_2}_{\left(\begin{smallmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 1 & 1 \end{smallmatrix}\right)}$. Recall that $\mathbf{1}^{\mathtt{X}_3}_{\left(\begin{smallmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 2 & 2 \end{smallmatrix}\right)}$ is the identity matrix where the column vector indexed by $\mathtt{X}_3$ is replaced by the vector $\begin{pmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 2 & 2 \end{pmatrix}$, that is

$$\mathbf{1}^{\mathtt{X}_3}_{\left(\begin{smallmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 2 & 2 \end{smallmatrix}\right)} \;=\; \begin{pmatrix} & \mathtt{X}_1 & \mathtt{X}_2 & \mathtt{X}_3 \\ \mathtt{X}_1 & 1 & 0 & 2 \\ \mathtt{X}_2 & 0 & 1 & 0 \\ \mathtt{X}_3 & 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{1}^{\mathtt{X}_2}_{\left(\begin{smallmatrix} \mathtt{X}_1 & \mathtt{X}_3 \\ 1 & 1 \end{smallmatrix}\right)} \;=\; \begin{pmatrix} & \mathtt{X}_1 & \mathtt{X}_2 & \mathtt{X}_3 \\ \mathtt{X}_1 & 1 & 0 & 0 \\ \mathtt{X}_2 & 0 & 1 & 0 \\ \mathtt{X}_3 & 0 & 1 & 1 \end{pmatrix}$$

and the product is

$$\begin{pmatrix} & \mathtt{X}_1 & \mathtt{X}_2 & \mathtt{X}_3 \\ \mathtt{X}_1 & 1 & 0 & 2 \\ \mathtt{X}_2 & 0 & 1 & 0 \\ \mathtt{X}_3 & 0 & 0 & 2 \end{pmatrix} \;\otimes\; \begin{pmatrix} & \mathtt{X}_1 & \mathtt{X}_2 & \mathtt{X}_3 \\ \mathtt{X}_1 & 1 & 0 & 0 \\ \mathtt{X}_2 & 0 & 1 & 0 \\ \mathtt{X}_3 & 0 & 1 & 1 \end{pmatrix} \;=\; \begin{pmatrix} & \mathtt{X}_1 & \mathtt{X}_2 & \mathtt{X}_3 \\ \mathtt{X}_1 & 1 & 2 & 2 \\ \mathtt{X}_2 & 0 & 1 & 0 \\ \mathtt{X}_3 & 0 & 2 & 2 \end{pmatrix}$$

Hence, we have

$$[\![\mathtt{X}_3 := \mathtt{mul}(\mathtt{sub}(\mathtt{X}_1,\mathtt{X}_2),\mathtt{X}_3)\,;\, \mathtt{X}_2 := \mathtt{max}(\mathtt{X}_2,\mathtt{X}_3)]\!] \;=\; \begin{pmatrix} & \mathtt{X}_1 & \mathtt{X}_2 & \mathtt{X}_3 \\ \mathtt{X}_1 & 1 & 2 & 2 \\ \mathtt{X}_2 & 0 & 1 & 0 \\ \mathtt{X}_3 & 0 & 2 & 2 \end{pmatrix}$$

What does this matrix tell us? Let x_1, x_2, x_3 be the numbers stored in $\mathtt{X}_1, \mathtt{X}_2, \mathtt{X}_3$ when the execution of the sequence begins, and let x_1', x_2', x_3' be the values stored in $\mathtt{X}_1$, $\mathtt{X}_2, \mathtt{X}_3$ at execution end. The certificate obtained above implies

- $x'_1 \leq x_1$
- $x'_2 \leq x_2 + p(x_1, x_3)$ for some polynomial p; in this case $x'_2 \leq x_2 + (x_1 \times x_3)$.
- $x'_3 \leq q(x_1, x_3)$ for some polynomial q; in this case $x'_3 \leq x_1 \times x_3$.

Example The bound on a value computed inside a loop may or may not depend on how many times the loop's body is executed. The loop correction operator adds the appropriate dependence on the iteration variable to the value bounds.

Consider the program $\texttt{loop Z}\ \{\ \texttt{X := add(X, Y)}\ \}$. Assume the input value of $\texttt{Z}$ is z. A bound on the output value of $\texttt{X}$ depends on z, as the value held by $\texttt{X}$ will be increased z times by the content of $\texttt{Y}$. A bound on the output value of $\texttt{Y}$ does not depend on z as $\texttt{Y}$ is not modified inside the loop. The loop correction operator ensures that this is reflected in the interpretation of $\texttt{loop Z}\ \{\ \texttt{X := add(X, Y)}\ \}$. We invite the reader to check that

$$[\![\texttt{loop Z}\ \{\ \texttt{X := add(X, Y)}\ \}]\!] \;=\; ([\![\texttt{X := add(X, Y)}]\!]^*)^{\downarrow \texttt{Z}} \;=$$
$$([\texttt{X} \mapsto \left(\begin{smallmatrix}\texttt{XY}\\12\end{smallmatrix}\right), \texttt{Y} \mapsto \left(\begin{smallmatrix}\texttt{Y}\\1\end{smallmatrix}\right)]^*)^{\downarrow \texttt{Z}} \;=\; [\texttt{X} \mapsto \left(\begin{smallmatrix}\texttt{XY}\\12\end{smallmatrix}\right), \texttt{Y} \mapsto \left(\begin{smallmatrix}\texttt{Y}\\1\end{smallmatrix}\right)]^{\downarrow \texttt{Z}} \;=\; [\texttt{X} \mapsto \left(\begin{smallmatrix}\texttt{XYZ}\\122\end{smallmatrix}\right), \texttt{Y} \mapsto \left(\begin{smallmatrix}\texttt{Y}\\1\end{smallmatrix}\right)]\ .$$

Example
We will study the sequence $\texttt{loop Y}\ \{\ s\ \}$ where s is $\texttt{X}_1\texttt{ := add}(\texttt{X}_1, \texttt{X}_2)\texttt{; X}_2\texttt{ := add}(\texttt{X}_2, \texttt{X}_3)$. It is left to the reader to verify that $[\![s]\!] = \begin{pmatrix} & \texttt{X}_1 & \texttt{X}_2 & \texttt{X}_3 & \texttt{Y} \\ \texttt{X}_1 & 1 & 0 & 0 & 0 \\ \texttt{X}_2 & 2 & 1 & 0 & 0 \\ \texttt{X}_3 & 0 & 2 & 1 & 0 \\ \texttt{Y} & 0 & 0 & 0 & 1 \end{pmatrix}$. Now, $[\![s]\!]^*$ exists with $[\![s]\!]^* = \begin{pmatrix} & \texttt{X}_1 & \texttt{X}_2 & \texttt{X}_3 & \texttt{Y} \\ \texttt{X}_1 & 1 & 0 & 0 & 0 \\ \texttt{X}_2 & 2 & 1 & 0 & 0 \\ \texttt{X}_3 & 4 & 2 & 1 & 0 \\ \texttt{Y} & 0 & 0 & 0 & 1 \end{pmatrix}$. Hence we obtain

$$[\![\texttt{loop Y}\ \{\ s\ \}]\!] \;=\; [\![s]\!]^{*\downarrow \texttt{Y}} \;=\; \begin{pmatrix} & \texttt{X}_1 & \texttt{X}_2 & \texttt{X}_3 & \texttt{Y} \\ \texttt{X}_1 & 1 & 0 & 0 & 0 \\ \texttt{X}_2 & 2 & 1 & 0 & 0 \\ \texttt{X}_3 & 4 & 2 & 1 & 0 \\ \texttt{Y} & 0 & 0 & 0 & 1 \end{pmatrix}^{\downarrow \texttt{Y}} \;=\; \begin{pmatrix} & \texttt{X}_1 & \texttt{X}_2 & \texttt{X}_3 & \texttt{Y} \\ \texttt{X}_1 & 1 & 0 & 0 & 0 \\ \texttt{X}_2 & 2 & 1 & 0 & 0 \\ \texttt{X}_3 & 4 & 2 & 1 & 0 \\ \texttt{Y} & 4 & 2 & 0 & 1 \end{pmatrix}\ .$$

Now, let x_1, x_2, x_3, y be the numbers stored in respectively $\texttt{X}_1, \texttt{X}_2, \texttt{X}_3, \texttt{Y}$ when the execution of $\texttt{loop Y}\ \{\ s\ \}$ starts; and let x'_1, x'_2, x'_3, y' be the numbers stored in $\texttt{X}_1, \texttt{X}_2, \texttt{X}_3, \texttt{Y}$ when the execution ends. The computed certificate guarantees that

- $x'_1 \leq x_1 + p(x_2, x_3, y)$ for some polynomial p, e.g. $x'_1 \leq x_1 + y \times x_2 + \frac{y(y-1)}{2} \times x_3$.
- $x'_2 \leq x_2 + q(x_3, y)$ for some polynomial q, e.g. $x'_2 \leq x_2 + y \times x_3$.
- $x'_3 \leq x_3$ and $y' \leq y$.

Example
Finally, we shall study the sequence $\texttt{loop Y}\ \{\ s\ \}$ where s is $\texttt{X}_1\texttt{ := add}(\texttt{X}_1, \texttt{X}_2)\texttt{;}$ $\texttt{X}_2\texttt{ := add}(\texttt{X}_2, \texttt{X}_2)$. One can easily calculate $[\![s]\!] = \begin{pmatrix} & \texttt{X}_1 & \texttt{X}_2 & \texttt{Y} \\ \texttt{X}_1 & 1 & 0 & 0 \\ \texttt{X}_2 & 2 & 2 & 0 \\ \texttt{Y} & 0 & 0 & 1 \end{pmatrix}$. This is the matrix B from Example 1, for which the closure B^* does not exist. Hence, $[\![\texttt{loop Y}\ \{\ s\ \}]\!]$ is undefined, and we cannot certify the program. This is as desired, since some of the output values of the program are not polynomially bounded in the input values. Namely $x'_2 = x_2 \times 2^y$.

Theorem 1. *Assume that* $[\![s]\!] = M$ *for some matrix* M*. Then, for* $i = 1, \ldots, n$*, there exists a polynomial* p_i *such that* $\vec{a}[s]\vec{b} \Rightarrow b_i \leq \max(\vec{u}) + p_i(\vec{v})$ *where* a_j *is in the list* $\vec{u}$ *iff* $M_{X_i}[X_j] = 1$*; and* a_j *is in the list* $\vec{v}$ *iff* $M_{X_i}[X_j] > 1$*.*

This is an adaptation of the result of Jones & Kristiansen [JK09].

Corollary 1. *If* $[\![s]\!]$ *is defined, then* s *is feasible.*

3.3 Comments and Comparisons with Related Work

The *mwp*-analysis presented in [JK09], can be seen as an abstract interpretation method where first-order imperative programs are interpreted as matrices over a finite semiring. The method presented above builds on the insights of *mwp*-analysis and is, in certain respects, an improvement of *mwp*-analysis:

- By interpreting operators as λ-expressions over vectors, we can easily include any operator in our programming language. We have included `add`, `mul`, `sub` and `max`, and it is straightforward to extend this list, e.g. we can include e.g. `div` (integer division) by the interpretation $[\![\texttt{div}]\!] = \lambda x \lambda y \binom{x}{1}$. In contrast, the original *mwp*-analysis only admits the operators `add` and `mul`.
- We provide a technique for dealing with constants. The *mwp*-analysis in [JK09] assumes there are no constants in the programs.
- We interpret programs as matrices over an infinite semiring, and our matrices contains more information than the finite ones in [JK09]. When an *mwp*-matrix is assigned to a program, we know that there exists polynomial upper bounds on the output values of the program; the present analysis provides bounds on the degrees these polynomial bounds.

These improvements of *mwp*-analysis are side effects of our effort to construct an interpretation method which is suitable for lifting to a higher order setting. The reader should also be aware that our method in certain respects is weaker than *mwp*-analysis, which again is weaker than the method introduced in [BAJK08]. This is acceptable since our motivation is not to capture as many first-order algorithms as possible. Rather, it is to develop a theory and book-keeping framework that can easily be lifted to higher orders.

Hopefully, lifting the interpretation presented in this section to higher orders will be a useful and educating exercise. Eventually, we might embark on more difficult projects, like e.g. lifting the analysis in [BAJK08].

4 Analysis of a Higher-Order Programming Language

4.1 The Higher-Order Programming Language

Syntax We will now extend our first order programming language to a higher order language. This language has variables of any type; variables of type ι hold natural numbers and variables of type σ, where $\sigma \neq \iota$, hold compound types.

Definition. *Types*, *Expressions* and *sequences* are defined by the following grammar:

$$
\begin{array}{lll}
\text{(Types)} & \ni \sigma, \tau ::= \iota \mid \sigma \otimes \tau \mid \sigma \to \tau \\
\text{(Variables)} & ::= \mathtt{X} \\
\text{(Constants)} & ::= k_n \text{ for each } k \in \mathbb{N} \\
\text{(Operators)} & ::= \mathtt{op} \\
\text{(Expressions)} \ni & e ::= \mathtt{X} \mid k_n \mid \mathtt{op}(e_1, e_2) \mid \mathtt{pair}(e_1, e_2) \mid \mathtt{fst}(e) \mid \mathtt{snd}(e) \mid \\
 & \qquad \mathtt{app}(e_1, e_2) \mid \{s\}_{\mathtt{X}} \mid \mathtt{proc}(\mathtt{X})\ e \\
\text{(Sequences)} & \ni s ::= \varepsilon \mid s_1 \,;\, s_2 \mid \mathtt{X} := e \mid \mathtt{loop}\ \mathtt{X}^{\iota}\ \{\, s \,\}
\end{array}
$$

As before, the set of operators is unspecified, but apply only to first order (i.e. with expressions of type ι). The expression $\{s\}_{\mathtt{X}}$ returns the value of $\mathtt{X}$ after executing the sequence s, while all other variables are left unmodified (i.e. local copies are created and discarded as needed). It is mostly used as the body of procedure definitions, thus leading to a call by value mechanism.

The high-order language extends the first order Loop language by using high-order variables. It is close to the extension done by Crolard & al. [CPV09]. It can be verified that the resulting language is an imperative variant of Gödel's System T. Our transformation is easier than the one by Crolard & al., because we do not require the lock-step simulation of System T.

Typing Expressions are typed in a straightforward way by a classical typing system. We leave out the typing rules here as they are quite obvious. Product types are used for typing pairs, arrow types for procedures, and so forth. The notation $\sigma_1, \sigma_2, \ldots, \sigma_n \to \tau$ is shorthand for $\sigma_1 \to (\sigma_2 \to (\ldots(\sigma_n \to \tau)\ldots))$. E.g., a function from $\mathbb{N}^3$ into $\mathbb{N}$ will be of the type $\iota, \iota, \iota \to \iota$ which is shorthand for the type $\iota \to (\iota \to (\iota \to \iota))$.

When needed, we will use superscript to denote the type of an expression. That is, e^{σ} is an expression of type σ. The variable controlling a loop must be of type ι. There are infinitely many variables of each type, so that we never run out of variables.

Semantics Programs are expressions of type ι whose only free variables are of type ι. They are evaluated by a standard call by value semantics depicted on Figure 1.

An *environment* env maps variables to expressions. Updating an environment is written $\mathrm{env}\{\mathtt{X} \mapsto e\}$, $\mathrm{env}\{\mathtt{X} \mapsto \emptyset\}$ is used to unbind $\mathtt{X}$ when building λ-abstraction for procedures. Operators are evaluated via a given semantics function $\overline{\mathtt{op}}^{\iota,\iota\to\iota}$. $\vdash$ evaluates an expression in a given environnement while $\Vdash$ evaluates a sequence. $\vdash^*$ is the reflexive transitive closure of $\vdash$. Rule (Proc) reduces inside the procedure's body, thus leading to a static binding of variables. Rule (Loop) expands the loop to $\mathrm{env}(\mathtt{X})$ copies of its body, so even if the loop counter were allowed to occur within the body, the behaviour would remain unchanged. The rules for dealing with pairs are the obvious ones and are omitted here.

$$\frac{}{\mathtt{X}, \mathrm{env} \vdash \mathrm{env}(\mathtt{X})}\ (\textsc{Var}) \qquad \frac{}{k_n, \mathrm{env} \vdash k_n}\ (\textsc{Cons}) \qquad \frac{e_1, \mathrm{env} \vdash^* e_1' \quad e_2, \mathrm{env} \vdash^* e_2'}{\mathtt{op}(e_1, e_2), \mathrm{env} \vdash \overline{\mathtt{op}}(e_1', e_2')}\ (\textsc{Op})$$

$$\frac{e, \mathrm{env}\{\mathtt{X} \mapsto \emptyset\} \vdash^* e'}{\mathtt{proc}(\mathtt{X})\ e, \mathrm{env} \vdash \mathtt{proc}(\mathtt{X})\ e'}\ (\textsc{Proc}) \qquad \frac{s, \mathrm{env} \Vdash \mathrm{env}'}{\{s\}_{\mathtt{X}}, env \vdash \mathrm{env}'(\mathtt{X})}\ (\textsc{Return})$$

$$\frac{e_1, \mathrm{env} \vdash^* \mathtt{proc}(\mathtt{X})\ e_1' \quad e_2, \mathrm{env} \vdash^* e_2' \quad e_1', \mathrm{env}\{\mathtt{X} \mapsto e_2'\} \vdash^* e'}{\mathtt{app}(e_1, e_2), \mathrm{env} \vdash e'}\ (\textsc{App})$$

$$\frac{}{\varepsilon, \mathrm{env} \Vdash \mathrm{env}}\ (\textsc{Empty}) \qquad \frac{s_1, \mathrm{env} \Vdash \mathrm{env}' \quad s_2, \mathrm{env}' \Vdash \mathrm{env}''}{s_1\,;\,s_2, \mathrm{env} \Vdash \mathrm{env}''}\ (\textsc{Comp})$$

$$\frac{s\,;\,\ldots\,;\,s, \mathrm{env} \Vdash \mathrm{env}'}{\mathtt{loop}\ \mathtt{X}\ \{\ s\ \}, \mathrm{env} \Vdash \mathrm{env}'}\ (\textsc{Loop}) \qquad \frac{e, \mathrm{env} \vdash^* e'}{\mathtt{X} := e, \mathrm{env} \Vdash \mathrm{env}\{\mathtt{X} \mapsto e'\}}\ (\textsc{Assign})$$

Fig. 1. Call by value semantics

4.2 Higher Order Vectors and Matrices

Vectors In order to separate the programs and program's expressions from their interpretation (as matrices, vectors, and algebraic expressions), we consider a base type κ for algebraic expressions.

We have first order indices $x_0^\kappa, x_1^\kappa, x_2^\kappa, \ldots$, and $\mathcal{I}^\kappa$ denotes the set of first order indices. Let $\mathcal{S}$ be any finite closed semiring. We define the arrow types $\alpha \to \beta$ and product types $\alpha \otimes \beta$ over the base type κ in the standard way. Letters early in the Greek alphabet $\alpha, \beta, \gamma \ldots$ t denote types over the base type κ whereas $\sigma, \tau, \ldots$ denote types over the base type ι. For any type α, we have a set of indices $\mathcal{I}^\alpha = \{x_0^\alpha, x_1^\alpha, x_2^\alpha, \ldots\}$. $\mathcal{I}$ denotes the set of all index sets.

Next, we lift the standard addition operator on vectors to an operator on algebraic expressions of higher types. We define the addition operator $+_\alpha$ over the type α recursively over the structure of the type α.

- $a +_\kappa b = a + b$ (standard addition of vectors over $\mathcal{S}$ and $\mathcal{I}^\kappa$)
- $c = a +_{\alpha\to\beta} b$ iff $c(i) = a(i) +_\beta b(i)$ for every $i \in \alpha$ (pointwise addition of functions)
- $(a_1, a_2)\ +_{\alpha\otimes\beta}\ (b_1, b_2) = (a_1 +_\alpha b_1,\ a_2 +_\beta b_2)$

We define the *higher order vector expressions* over $\mathcal{S}$ and the index set $\mathcal{I}$ as the set containing the first order vectors (indexed by $\mathcal{I}^\kappa$), variables (i.e. indexes of $\mathcal{I}$) and which is closed by addition, λ-abstraction, application and pairing. If e^β is an expression and x^α is a free index in e, then $e[x \setminus t^\alpha]$ denotes the

expression resulting from replacing all free occurrences of x in e by t. To improve the readability, we will write $e[x_1 \setminus t_1, \ldots, x_n \setminus t_n]$ in place of $e[x_1 \setminus t_1] \ldots [x_n \setminus t_n]$.

Matrices A *higher order matrix* over $\mathcal{S}$ and $\mathcal{I}$ is a mapping from $\mathcal{I}$ into the higher order vector expressions over $\mathcal{S}$ and $\mathcal{I}$ such that each index of type α is mapped to an expression of type α. We use $M, A, B, C, \ldots$ to denote higher order matrices, and M_x denotes the expression to which M maps the index x. The *unity matrix* matrix $\mathbf{1}$ is the matrix such that $\mathbf{1}_x = \binom{x}{1}$ when $x \in \mathcal{I}^\kappa$ is an index of the base type κ; and $\mathbf{1}_x = x$ when $x \in \mathcal{I} \setminus \mathcal{I}^\kappa$. The *sum* $A{+}B$ of the matrices A and B is defined by

$$M = A{+}B \quad \Leftrightarrow \quad M_x = A_x +_\sigma B_x$$

for any index x of type σ. The matrix A is an *upper bound* of the matrix B, written $A \geq B$, if there exists a matrix C such that $A = B{+}C$. Thus we have a partial ordering of the higher order matrices. The ordering symbols $\geq, \leq, >, <$ have their standard meaning with respect to this ordering. The *product* $A{\cdot}B$ of the matrices A and B is defined by

$$M = A{\cdot}B \quad \Leftrightarrow \quad M_x = B_x[y_1 \setminus A_{y_1}, \ldots, y_n \setminus A_{y_n}]$$

where $y_1, \ldots, y_n$ are the free indices in B_x.

It is worth noticing that this multiplication extends the usual matrix multiplication in the sense that the product, as defined above, of two first-order matrices (i.e. whose high order indexes are mapped to 0) is the same as the usual matrix product.

We can now lift the closure operator defined in Section 2 to higher order matrices. We define the closure M^* of the higher order matrix M by

$$M^* \quad = \quad \mathbf{1} + M^1 + M^2 + M^3 + \ldots$$

where $M^1 = M$ and $M^{n+1} = M{\cdot}M^n$.

We will use the sparse representation $\begin{pmatrix} X_1 \mapsto e_1 \\ \vdots \\ X_n \mapsto e_n \end{pmatrix}$ to denote the higher order matrix where the index X_1 maps to the vector expression e_1, the index X_2 maps to the vector expression e_2, and so on. Furthermore, any index Y not occurring in the list $X_1, \ldots, X_n$ maps to Y.

4.3 Interpretation of Programs

Let $\mathcal{S}$ be the finite closed semiring $(\mathbb{N}, \oplus, \otimes, 0, 1)$ defined in Section 3.2. We map the types over ι to the types over κ by $\overline{\iota} = \kappa$, $\overline{\sigma \to \tau} = \overline{\sigma} \to \overline{\tau}$, and $\overline{\sigma \otimes \tau} = \overline{\sigma} \otimes \overline{\tau}$. Let the index set $\mathcal{I}$ be the set of program variables, any variable $\mathtt{X} \in \mathcal{I}^\sigma$ will serve as an index of type $\overline{\sigma}$. We will interpret program expressions and program sequences as respectively vector expressions and matrices over $\mathcal{S}$ and $\mathcal{I}$. Before we can define the interpretation operator $[\![\cdot]\!]$, we need to lift the loop correction

operator from Section 3.2 to higher order matrices.

The function $\Phi^\alpha : \mathcal{I}^\kappa \times \alpha \longrightarrow \alpha$ is given by

- $\Phi^\kappa(x, V) = \begin{cases} V \oplus \binom{x}{a} \text{ if } a > 1 & \text{where } a = \sum_{y \in \mathcal{I}^\kappa} V[y] \\ V & \text{otherwise} \end{cases}$
- $\Phi^{\alpha \to \beta}(x, W) = \lambda U^\alpha \Phi^\beta(x, W(U))$
- $\Phi^{\alpha \otimes \beta}(x, W) = \langle \Phi^\alpha(x, W_0), \Phi^\beta(x, W_1) \rangle$, where W_0 and W_1 respectively denote the left and the right component of the pair W

When M is a high order matrix, then $M^{\downarrow x}$ is the matrix A where $A_y = \Phi^\alpha(x, M_y)$ for each $y^\alpha \in \mathcal{I}$. We say that the loop correction $M^{\downarrow x}$ takes place with respect to the first order index x. Next, we define the interpretation operator $[\![\cdot]\!]$.

- Interpretations of expressions:
  - for any program variable $\mathtt{X}^\iota$ let $[\![\mathtt{X}]\!] = \binom{\mathtt{X}}{1}$
  - for any program variable $\mathtt{X}^\sigma$ where $\sigma \neq \iota$ let $[\![\mathtt{X}]\!] = x^{\overline{\sigma}}$
  - $[\![\mathtt{proc}(\mathtt{X})\, e]\!] = \lambda x [\![e]\!]$
  - $[\![\mathtt{app}(e_1, e_2)]\!] = [\![e_1]\!]([\![e_2]\!])$
  - $[\![\mathtt{pair}(e_1, e_2)]\!] = \langle [\![e_1]\!], [\![e_1]\!] \rangle$
  - $[\![\mathtt{fst}(e)]\!] = \ell$ where ℓ is the first component of the pair $[\![e]\!]$
  - $[\![\mathtt{snd}(e)]\!] = \ell$ where ℓ is the second component of the pair $[\![e]\!]$
  - $[\![\mathtt{add}]\!] = \lambda x^\kappa \lambda y^\kappa \binom{x\ y}{1\ 2}$ $\quad [\![\mathtt{mul}]\!] = \lambda x \lambda y \binom{x\ y}{2\ 2}$
  - $[\![\mathtt{sub}]\!] = \lambda x \lambda y \binom{x}{1}$ $\quad [\![\mathtt{max}]\!] = \lambda x \lambda y \binom{x\ y}{1\ 1}$
  - $[\![\{s\}_\mathtt{X}]\!] = [\![s]\!]_\mathtt{X}$
- Interpretations of sequences:
  - $[\![s_1 \,;\, s_2]\!] = [\![s_1]\!] \cdot [\![s_2]\!]$
  - $[\![x := e]\!] = (\mathbf{1}^x_{[\![e]\!]})$ where $\mathbf{1}^x_{[\![e]\!]}$ denotes the identity matrix $\mathbf{1}$ where the vector indexed by x is replaced by the vector $[\![e]\!]$
  - $[\![\varepsilon]\!] = \mathbf{1}$ (the identity matrix)
  - $[\![\mathtt{loop}\ \mathtt{X}\ \{\ s\ \}]\!] = \begin{cases} ([\![s]\!]^*)^{\downarrow \mathtt{X}} & \text{if } ([\![s]\!]^*)_i[i] = 1 \text{ for all } i \in \mathcal{I}^\kappa \\ \text{undefined otherwise} \end{cases}$

5 Conjectures

A sequence s is a *program* iff, for some $n > 0$, the first order variables $\mathtt{X}^\iota_1, \ldots, \mathtt{X}^\iota_n$ are the only free variables occurring in s.

The program execution relation $a_1, \ldots, a_n[s]b_1, \ldots, b_n$ holds if and only if the variables $\mathtt{X}_1, \ldots, \mathtt{X}_n$ respectively hold the numbers $a_1, \ldots, a_n$ when the execution of s starts and the numbers $b_1, \ldots, b_n$ when the execution terminates; that is, the program execution relation is defined exactly as for first-order programs.

Conjecture 1. Let $2^x_0 = x$ and $2^x_{\ell+1} = 2^{2^x_\ell}$, and let s be a program. Assume that there exists a matrix M such that $[\![s]\!] = M$. Then, for $i = 1, \ldots, n$, there exists $\ell \in \mathbb{N}$ such that

$$\vec{a}[s]\vec{b} \ \Rightarrow b_i \ \leq \max(\vec{u}) + 2^{\max(\vec{v},1)}_\ell$$

where a_j is in the list $\vec{u}$ iff $M_{\mathtt{X}_i}[\mathtt{X}_j] = 1$; and a_j is in the list $\vec{v}$ iff $M_{\mathtt{X}_i}[\mathtt{X}_j] \geq 2$.

Conjecture 2. The existence of M with $[\![s]\!] = M$ is decidable.

We will now give examples that justify the conjectures. However, no proofs are provided.

6 Justifying the Conjectures

In our higher order language we can write a sequence p of the form

$$\begin{aligned}&\mathtt{F} := \mathtt{proc}(\mathtt{f}^{\iota\to\iota})\,\mathtt{proc}(\mathtt{Z}^{\iota})\\ &\qquad\{\mathtt{loop\ Z}\,\{\,\mathtt{f} := \mathtt{proc}(\mathtt{U}^{\iota})\,\mathtt{ff}(\mathtt{U})\,\}\}_{\mathtt{f}}\,;\\ &\mathtt{X}^{\iota} := \mathtt{F}(\mathtt{g},\mathtt{Y})(\mathtt{Z})\end{aligned}$$

where $\mathtt{g}$ is some expression of type $\iota \to \iota$, that is, a function $\mathbb{N} \to \mathbb{N}$. To improve the readability, we will write $\mathtt{f}(\mathtt{X})$ in place of $\mathtt{app}(\mathtt{f},\mathtt{X})$, $\mathtt{F}(\mathtt{g},\mathtt{Y})$ in place of $\mathtt{app}(\mathtt{app}(\mathtt{F},\mathtt{g}),\mathtt{Y})$, et cetera. The sequence assigns a functional to $\mathtt{F}$ such that $\mathtt{F}(\mathtt{g},n)(\mathtt{X}) = \mathtt{g}^{2^n}(\mathtt{X})$, that is, the expression $\mathtt{F}(\mathtt{g},k_n)$ executes the function $\mathtt{g}$ repeatedly 2^n times on its input argument and returns the result. Thus, the last command of the sequence assigns to the numeric variable $\mathtt{X}$ the result of applying a function over and over again, $2^{\mathtt{Y}}$ times, to the value held by $\mathtt{Z}$. Let s_F be the sequence defining $\mathtt{F}$ (the right hand side of the first assignement), we have

$$[\![s_F]\!] \;=\; \lambda f\lambda z((\,[f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^{*})^{\downarrow z})_f$$

Note that we used typewriter font for the program variables while we now use normal font for the high-order algebraic expression variables that interpret them. Now we have:

$$\lambda u f f\,(\,{}^{u}_{1}\,)\,[f \setminus \lambda u f f\,(\,{}^{u}_{1}\,)] \;=\; \lambda u(\lambda u f f\,(\,{}^{u}_{1}\,)\,(\lambda u f f\,(\,{}^{u}_{1}\,)\,(\,{}^{u}_{1}\,))) \;=\; \lambda u f f f f\,(\,{}^{u}_{1}\,)$$

Hence,

$$[f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^2 \;=\; [f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)] \cdot [f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)] \;=\; [f \mapsto \lambda u f f f f\,(\,{}^{u}_{1}\,)]\,.$$

By the same token we have

$$\begin{aligned}[f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^3 &= [f \mapsto \lambda u f f f f f f f f f\,(\,{}^{u}_{1}\,)]\\ [f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^4 &= [f \mapsto \lambda u f f f f f f f f f f f f f f f f\,(\,{}^{u}_{1}\,)]\\ &\qquad\text{etc.}\end{aligned}$$

and thus

$$\begin{aligned}[f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^* \;=\; \mathbf{1} +_\sigma [f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)] +_\sigma [f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^2 +_\sigma \ldots \;=\\ [f \mapsto \mathbf{1} +_\sigma \lambda u f f\,(\,{}^{u}_{1}\,) +_\sigma \lambda u f f f f\,(\,{}^{u}_{1}\,) +_\sigma \ldots]\end{aligned}$$

where $\sigma \equiv \kappa \to \kappa$. If the infinite sum on the right hand side converges to the value A, then

$$[f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^* \;=\; [f \mapsto A]$$

Now, A is an expression of type $\kappa \to \kappa$, and

$$[f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^{*\downarrow z} \;=\; [f \mapsto A]^{\downarrow z} \;=\; [f \mapsto \Phi^{\iota\to\iota}(z, A)] \;=\; [f \mapsto \lambda u \Phi^{\iota}(z, Au)]$$

The two last equalities hold by the definition of the loop correction operator. Notice that we are not yet able to compute the actual value of this interpretation. Indeed, we first need to instantiate the λ-abstraction within it before computing. So it is not possible, at this point, to decide, e.g., whether the closure exists. This is expected, as it is the interpretation of the functional `F`, whose behaviour (with respect to the bound on the computed value) depends not only on the first order values, but also on its high-order arguments (namely `f`, being interpreted as the abstracted variable f). Now, we see that

$$[\![s_F]\!] \quad = \quad \lambda f \lambda z((([f \;\mapsto\; \lambda u f f\,(\,{}^{u}_{1}\,)]^{*})^{\downarrow z})_f \quad = \quad \lambda f \lambda z(\lambda u \Phi^{\iota}(z, Au)) \quad (*)$$

where $A \equiv 1_\sigma +_\sigma \lambda u f f\,(\,{}^{u}_{1}\,) +_\sigma \lambda u f f f f\,(\,{}^{u}_{1}\,) +_\sigma \ldots$ and

$$\Phi^{\iota}(z, Au) \;=\; \begin{cases} Au \oplus (\,{}^{z}_{2}\,) & \text{if } \sum_{y\in\mathcal{I}}(Au)[y] > 1 \\ Au & \text{otherwise} \end{cases}$$

What happens now, if the `g` in our program is the identity function, and what happens if `g` is the doubling function? First, assume that `g` is the identity function, i.e. we have the program

$$s_0 \;=\; \begin{array}{l} \texttt{F:=proc(f}^{\iota\to\iota}\texttt{)\,proc(Z}^{\iota}\texttt{)\{loop Z\,\{\,f:=proc(U}^{\iota}\texttt{)\,ff(U)\,\}\}}_{\texttt{f}}\texttt{;} \\ \texttt{X}^{\iota}\texttt{:=F(proc(X)X, Y)(Z)} \end{array}$$

Then,

$$[\![s_0]\!] \;\equiv\; \begin{pmatrix} \texttt{F} \mapsto \lambda f \lambda z((([f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^{*})^{\downarrow z})_f \\ \texttt{X} \mapsto \lambda f \lambda z((([f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^{*})^{\downarrow z})_f([\![\lambda x.x]\!], [\![y]\!])[\![z]\!] \end{pmatrix}$$

If f is the identity function, that is the function $\lambda x^{\kappa}\,(\,{}^{x}_{1}\,)$, then get

$$\begin{array}{l} A \equiv 1_\sigma +_\sigma \lambda u f f\,(\,{}^{u}_{1}\,) +_\sigma \lambda u f f f f\,(\,{}^{u}_{1}\,) +_\sigma \ldots \;=\; \lambda u \lambda x^{\iota}\,(\,{}^{x}_{1}\,)\,(\lambda x^{\iota}\,(\,{}^{x}_{1}\,)\,(\,{}^{u}_{1}\,) \;+_\sigma \\ \qquad \lambda u \lambda x^{\iota}\,(\,{}^{x}_{1}\,)\,(\lambda x^{\iota}\,(\,{}^{x}_{1}\,)\,(\lambda x^{\iota}\,(\,{}^{x}_{1}\,)\,(\lambda x^{\iota}\,(\,{}^{x}_{1}\,)\,(\,{}^{u}_{1}\,))))\;+_\sigma\; \ldots \;=\; \lambda u\,(\,{}^{u}_{1}\,)\;. \end{array}$$

Hence, by (*) we obtain

$$\begin{array}{l} \lambda f \lambda z((([f \mapsto \lambda u f f\,(\,{}^{u}_{1}\,)]^{*})^{\downarrow z})_f([\![\texttt{proc(X)}\;\{\}_{\texttt{X}}]\!], [\![\texttt{Y}]\!]) \;= \\ \quad \lambda f \lambda z(\lambda u \Phi^{\iota}(z, Au)(\lambda x^{\iota}\,(\,{}^{x}_{1}\,), (\,{}^{y}_{1}\,)) \;=\; \lambda z(\lambda u \Phi^{\iota}(z, \lambda u\,(\,{}^{u}_{1}\,)\,u))\,(\,{}^{y}_{1}\,) \;= \\ \qquad \lambda u \Phi^{\iota}((\,{}^{y}_{1}\,), (\,{}^{u}_{1}\,)) \;=\; \lambda u \begin{cases} (\,{}^{u}_{1}\,) \oplus (\,{}^{y}_{2}\,) & \text{if } \sum_{i\in\mathcal{I}}(\,{}^{u}_{1}\,)\,[i] > 1 \\ (\,{}^{u}_{1}\,) & \text{otherwise} \end{cases} \;=\; \lambda u\,(\,{}^{u}_{1}\,) \end{array}$$

Finally, we get

$$[\![s_0]\!]_{\texttt{X}} \;=\; ((\lambda f \lambda z(\lambda u \Phi^{\iota}(z, Au))\lambda x^{\iota}\,(\,{}^{x}_{1}\,))\,(\,{}^{\texttt{Y}}_{1}\,)\,(\,{}^{\texttt{Z}}_{1}\,) \;=\; \lambda u\,(\,{}^{u}_{1}\,)\,(\,{}^{\texttt{Z}}_{1}\,) \;=\; \binom{\texttt{Z}}{1}\;.$$

Thus, the value assigned to the first order index $\mathtt{X}$ by the interpretation $[\![s_0]\!]$ is the vector $\binom{\mathtt{Z}}{1}$. This is good. If we inspect the program, we see that if a is held by $\mathtt{Z}$ when the execution of s_0 starts, and b is held by $\mathtt{X}$ when the execution terminates, then $b \leq a$.

What, then, will the interpretation of our program look like if we call the functional $\mathtt{F}$ with the doubling function instead of the identity function? That is, if we have the program

$$s_1 \equiv \begin{array}{l} \mathtt{F} := \mathtt{proc}(\mathtt{f}^{\iota\to\iota})\,\mathtt{proc}(\mathtt{Z}^{\iota})\{\mathtt{loop}\ \mathtt{Z}\ \{\,\mathtt{f} := \mathtt{proc}(\mathtt{U}^{\iota})\,\mathtt{ff}(\mathtt{U})\,\}_{\mathtt{f}}\,; \\ \mathtt{X}^{\iota} := \mathtt{F}(\mathtt{add},\mathtt{Y})(\mathtt{Z})\ . \end{array}$$

The answer is: The interpretation $[\![s_1]\!]$ does not exist.

Note that $[\![\mathtt{proc}(\mathtt{X})\ \{\mathtt{X} := \mathtt{add}(\mathtt{X},\mathtt{X})\}_{\mathtt{X}}]\!] = \lambda x\binom{x}{2}$. Thus, computing $[\![s_1]\!]$ involves computing the infinite sum

$$1_\sigma +_\sigma \lambda u f f\binom{u}{1} +_\sigma \lambda u f f f f\binom{u}{1} +_\sigma \lambda u f f f f f f f f\binom{u}{1} +_\sigma \ldots$$

where f is the function $\lambda x\binom{x}{2}$. We get

$$\begin{array}{l} 1_\sigma\ +_\sigma\ \lambda u \lambda x\binom{x}{2}\left(\lambda x\binom{x}{2}\binom{u}{1}\right) +_\sigma \\ \qquad \lambda u \lambda x\binom{x}{2}\left(\lambda x\binom{x}{2}\left(\lambda x\binom{x}{2}\left(\lambda x\binom{x}{2}\binom{u}{1}\right)\right)\right)\ +_\sigma\ \ldots\ = \\ \qquad\qquad \lambda u\binom{u}{4}\ +_\sigma\ \lambda u\binom{u}{16}\ +_\sigma\ \lambda u\binom{u}{256}\ +_\sigma\ \ldots \end{array}$$

This sum does not converge, and hence, the interpretation of program s_1 is undefined.

Now, let us show both the use of the constants k_n and an example where the program is certified, even if the computed values are not polynomial. Let $\mathtt{g}$ in the program above be the successor function. That is, we have the program s_2 where

$$s_2 \equiv \begin{array}{l} \mathtt{F} := \mathtt{proc}(\mathtt{f}^{\iota\to\iota})\,\mathtt{proc}(\mathtt{Z}^{\iota})\{\mathtt{loop}\ \mathtt{Z}\ \{\,\mathtt{f} := \mathtt{proc}(\mathtt{U}^{\iota})\,\mathtt{ff}(\mathtt{U})\,\}_{\mathtt{f}}\,; \\ S^{\iota\to\iota} := \mathtt{proc}(\mathtt{X})\{\mathtt{add}(\mathtt{X},k_1)\}_{\mathtt{X}}\,;\ \mathtt{X}^{\iota} := \mathtt{F}(S,\mathtt{Y})(\mathtt{Z})\ . \end{array}$$

This program computes the value $\mathtt{Z} + 2^{\mathtt{Y}}$ and stores it in $\mathtt{X}$, and the program has a certificate. Notice first that $[\![S]\!] = \lambda x\binom{x\ k}{1\ 2}$, where k is the special index dedicated to the constants. Computing $[\![s_2]\!]$ involves computing the infinite sum

$$1_\sigma +_\sigma \lambda u f f\binom{u}{1} +_\sigma \lambda u f f f f\binom{u}{1} +_\sigma \lambda u f f f f f f f f\binom{u}{1} +_\sigma \ldots$$

where f is the function $\lambda x\binom{x\ k}{1\ 2}$. This sum converges and equals $\lambda u\binom{u\ k}{1\ 2}$. This entails that the program has an interpretation. This certificate has $x \mapsto \binom{z\ y\ k}{1\ 2\ 2}$.

References

[BAJK08] A.M Ben-Amram, N.D. Jones, and L. Kristiansen. Linear, polynomial or exponential? Complexity inference in polynomial time. In *CiE'08:Logic and Theory of Algorithms*, volume 5028 of *LNCS*, pages 67–76. Springer, 2008.

[BMM07] G. Bonfante, J.-Y. Marion, and J.-Y. Moyen. Quasi-interpretations – a way to control resources. *Theoretical Computer Science*, 2007. To appear.

[MP09] J.-Y. Marion, and R. Pechoux. Sup-interpretations, a semantic method for static analysis of program resources. *AMS Transactions on Coputational Logic* 10, 2009.

[B04] R. Benzinger. Automated higher-order complexity analysis. *Theoretical Computer Science* 318 (2009), 79-103.

[CPV09] T. Crolard, E. Polonowski, and P. Valarcher. Extending the loop language with higher-order procedural variables. *Special issue of ACM TOCL on Implicit Computational Complexity*, 10(4):1–36, 2009.

[JK09] N. D. Jones and L. Kristiansen. A flow calculus of mwp-bounds for complexity analysis. *ACM Transactions of Computational Logic*, 10, 2009.

[KJ05] L. Kristiansen and N.D. Jones. The flow of data and the complexity of algorithms. In Torenvliet (eds.) Cooper, Lwe, editor, *CiE'05:New Computational Paradigms*, volume 3526 of *LNCS*, pages 263–274. Springer, 2005.

[MR76] A. R. Meyer and D. M. Ritchie. The complexity of loop programs. In *Proc. ACM Nat. Meeting*, 1976.

[NW06] K.-H. Niggl and H. Wunderlich. Certifying polynomial time and linear/polynomial space for imperative programs. *SIAM Journal on Computing*, 35(5):1122–1147, March 2006. published electronically.