

COMPUTATIONAL CHEMISTRY IN PARALLEL

by

**Dylan Hennessey**

A thesis submitted in partial fulfillment of the requirements for the degree of

**Master of Science.**

Department of Chemistry  
**University of Alberta**

© Dylan Hennessey, 2018

# Abstract

put abstract here

# Preface

put preface here

put dedication here

# Acknowledgements

Put acknowledgements here

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Solving the Schrödinger Equation . . . . .	1
1.2	Hartree-Fock . . . . .	4
1.2.1	Two-Electron Integrals . . . . .	5
1.2.2	Basis Sets . . . . .	7
1.2.3	Hartree-Fock SCF . . . . .	8
1.3	Special Relativity . . . . .	10
1.3.1	Relativistic Quantum Mechanics . . . . .	11
1.3.2	Dirac Equation . . . . .	12
1.3.3	Kinetic Balancing . . . . .	14
1.3.4	The Dirac Hydrogen-Like Atom . . . . .	15
1.3.5	Relativistic Effects . . . . .	15
1.3.6	Gaussian Nucleus . . . . .	16
<b>2</b>	<b>cudaDFRATOM</b>	<b>17</b>
2.1	GPU Architecture . . . . .	17
2.1.1	Threads, Blocks, and Grids . . . . .	17
2.1.2	Memory . . . . .	18
2.1.3	Occupancy and Performance . . . . .	20
2.1.4	Matrix Multiplication Example . . . . .	22
2.2	Program flow . . . . .	25
2.3	Alterations for CUDA . . . . .	26
2.3.1	Two-electron Integrals . . . . .	26
2.3.2	Density Matrix Formation . . . . .	37
2.4	Discussion . . . . .	44
2.4.1	Resource Usage . . . . .	46
2.4.2	Two-Electron Integral performance . . . . .	47
2.4.3	P and Q Matrix Formation Performance . . . . .	48
2.4.4	Total Speedup . . . . .	53

2.5	Input Description . . . . .	55
<b>3</b>	<b>Basis Sets</b>	<b>58</b>
3.1	Introduction . . . . .	58
3.2	Methods . . . . .	60
3.2.1	Basis set size optimization . . . . .	60
3.2.2	Automating Optimization . . . . .	62
3.3	Discussion . . . . .	68
3.3.1	Optimized Non-Relativistic Basis Sets . . . . .	68
	<b>Bibliography</b>	<b>76</b>

# List of Tables

1.1	The first few exact solutions for the hydrogen atom wavefunction. . . . .	3
2.1	Kernel resource usage . . . . .	46
2.2	<i>eint2</i> kernel performance . . . . .	47
2.3	<i>formpq_alg1</i> kernel profiling results. . . . .	49
2.4	<i>formpq_alg2</i> kernel profiling results. . . . .	49
2.5	Total speedup of alg1 . . . . .	53
2.6	Total speedup of alg2 . . . . .	53
3.1	Basis sets optimized using rwtbs . . . . .	68
3.2	Basis sets optimized using cudaDFRATOM . . . . .	70



# List of Figures

2.1	Architecture of a GPU . . . . .	21
2.2	Two-electron integral layout: naive implementation. . . . .	30
2.3	Two-electron integral layout: 2D blocks and grid . . . . .	31
2.4	Two-electron integral layout: 1D blocks and grid . . . . .	32
2.5	Two-electron integral matrix as a matrix . . . . .	34
2.6	Two-electron integral matrix as a vector . . . . .	35
2.7	Density matrix as a matrix . . . . .	38
2.8	Density matrix as a vector . . . . .	39
2.9	Profiling of two-electron integral calculations . . . . .	48
2.10	Profiling of the <i>formpq</i> kernels. . . . .	50
2.11	Integrals needed for the first element of $\mathbf{P}$ : matrix representation . . . . .	51
2.12	Integrals needed for the first element of $\mathbf{P}$ : vector representation . . . . .	52
2.13	Total time need to complete a calculation. . . . .	54
3.1	Changes in $\zeta$ as the values of $\delta$ and $\gamma$ are changed. . . . .	60
3.2	Scatter plot of the $\alpha$ parameter vs. nuclear charge. . . . .	72
3.3	Scatter plot of the $\beta$ parameter vs. nuclear charge. . . . .	73
3.4	Scatter plots of the $\delta$ and $\gamma$ parameters. . . . .	73
3.5	Scatter plot of the relativistic and non-relativistic $\alpha$ parameters vs. nuclear charge. . . . .	74
3.6	Scatter plot of the relativistic and non-relativistic $\beta$ parameters vs. nuclear charge. . . . .	75
3.7	Scatter plot of the relativistic and non-relativistic $\delta$ and $\gamma$ parameters. . . . .	75

# Chapter 1

## Introduction

### 1.1 Solving the Schrödinger Equation

The Schrödinger equation[1] in its time-independent, non-relativistic form is given in Equation 1.1

$$\hat{H}\Psi = E\Psi \quad (1.1)$$

where  $\hat{H}$  is the Hamiltonian operator,  $\Psi$  is referred to as the wavefunction, and  $E$  is the energy of the system described by the wavefunction. For a molecular system, the Hamiltonian operator in atomic units (a.u) is given by Equation 1.2

$$\hat{H} = -\frac{1}{2} \sum_i^n \nabla_i^2 - \sum_A^N \frac{1}{2M_A} \nabla_A^2 - \sum_i^n \sum_A^N \frac{Z_A}{r_{iA}} + \sum_i^n \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_A^N \sum_{B>A}^N \frac{Z_A Z_B}{r_{AB}} \quad (1.2)$$

where  $i$  and  $j$  refer to electrons,  $A$  and  $B$  refer to nuclei, and  $n$  and  $N$  refer to the maximum number of electrons and nuclei respectively.  $M_A$  and  $Z_A$  are the mass and charge of nucleus  $A$ .  $r_{xy}$  is the distance between particles  $x$  and  $y$ , and  $\nabla_n^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$  where  $x$ ,  $y$ , and  $z$  are the Cartesian coordinates of particle  $n$ . Each term accounts for a different kind of energy:  $-\frac{1}{2} \sum_i^n \nabla_i^2$  is the kinetic energy of the electrons,  $\sum_A^N \frac{1}{2M_A} \nabla_A^2$  is the kinetic energy of the nuclei,  $\sum_i^n \sum_A^N \frac{Z_A}{r_{iA}}$  is the potential energy of attraction between nuclei and electrons,  $\sum_i^n \sum_{j>i}^N \frac{1}{r_{ij}}$  is the potential energy of repulsion between electrons, and  $\sum_A^N \sum_{B>A}^N \frac{Z_A Z_B}{r_{AB}}$  is the potential energy of repulsion between nuclei. Together, this accounts for all of the energy of a molecular system. As this operator is quite complex, it would be helpful to simplify it if possible. Fortunately, Born and Oppenheimer[2] realized that an approximation could be made that would substantially reduce the amount of work we have to do.

Much like a satellite in orbit around the Earth, the proton and neutron vastly outweigh the electron. If we look at the second term in Equation 1.2, we notice that it is inversely

proportional to the mass of the nuclei. We can therefore assume, to a decent approximation, that the kinetic energy of the nuclei is very close to zero compared to the other terms. This allows us to treat them as stationary classical particles. As a bonus, this turns the potential energy of repulsion between nuclei into a constant. The killing of these two terms reduces the Hamiltonian operator to the *electronic* Hamiltonian given in Equation 1.3

$$\hat{H}_{el} = -\frac{1}{2} \sum_i^n \nabla_i^2 - \sum_i^n \sum_A^N \frac{Z_A}{r_{iA}} + \sum_i^n \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{B>A}^N \frac{Z_A Z_B}{r_{AB}} \quad (1.3)$$

$\hat{H}$  can be assumed to be  $\hat{H}_{el}$  for the remainder of this Thesis, and with the Hamiltonian out of the way, we can now turn our attention to the wavefunction. While we were able to get away with turning nuclei into classical particles, the electrons unfortunately remain fully quantum. Therefore, we need to deal with all the complicated mathematical baggage this comes with.

Consider the hydrogen atom which, like all other one-electron atomic systems, is one of the few exactly solvable problems in quantum mechanics. The solution for its wavefunction in spherical coordinates is given in Equation 1.4[3]

$$\Psi_{nlm}(r, \theta, \varphi) = R_{nl}(r) Y_l^m(\theta, \varphi) \quad (1.4)$$

where  $R_{nl}$  is called the radial function and  $Y_l^m(\theta, \varphi)$  is called the angular function. The radial equation has the form

$$R_{nl}(r) = \frac{1}{r} \left( \frac{r}{a_o n} \right)^{l+1} e^{-\frac{r}{a_o n}} L_{n-l-1}^{2l+1} \left( \frac{r}{a_o n} \right) \quad (1.5)$$

where  $a_o$  is the Bohr radius and  $L_{n-l-1}^{2l+1} \left( \frac{r}{a_o n} \right)$  is polynomial of  $\left( \frac{r}{a_o n} \right)$  and order  $j_{\max} = n - l - 1$  of  $\left( \frac{r}{a_o n} \right)$ . Its coefficients are given by

$$c_{j+1} = \frac{2(j+l+1-n)}{(j+1)(j+2l+2)} c_j \quad (1.6)$$

and the polynomial itself given by

$$L_{q-p}^p(x) = (-1)^p \left( \frac{d}{dx} \right)^q L_q(x) \quad (1.7)$$

$$L_q(x) = e^x \left( \frac{d}{dx} \right)^q [x^q e^{-x}] \quad (1.8)$$

The angular function has the form

$$Y_l^m(\theta, \varphi) = \epsilon \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}} P_l^m(\cos\theta) e^{im\varphi} \quad (1.9)$$

where  $i = \sqrt{-1}$  and  $\epsilon = (-1)^m$  for  $m \geq 0$  or 1 if  $m < 0$ .  $P_l^m(\cos\theta)$  is a polynomial of  $\cos\theta$  and has the form of

$$P_l^m(x) = (1-x^2)^{\frac{|m|}{2}} \frac{d^{|m|}}{dx^{|m|}} P_l(x) \quad (1.10)$$

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l \quad (1.11)$$

Inspection of these equations reveals that  $n$  must be a positive integer greater than 0,  $l$  must be a positive integer between 0 and  $n-1$ , and  $|m|$  can be no larger than  $l$ . These are called quantum numbers, specifically  $n$  is the principal quantum number,  $l$  is the azimuthal quantum number, and  $m$  is the magnetic quantum number. Also notice that there is no limit to how large  $n$  can get, which suggests that there is not just one solution to this equation, but an infinite number of them. The solutions to the hydrogen-like atom wavefunction are given the special name of orbitals. Also, orbitals with the same value of  $l$  are referred to be of the same symmetry. Table 1.1 shows the reduced form of  $\Psi$  for a few different orbitals. For atoms with more than one electron, if we make a further assumption that electrons can be modelled like ideal springs, it has been shown that an atomic system with two electrons can also be solved[4], but we can go no further than this. At least, not exactly. In the next section we will see how solutions for atoms, and indeed molecules, can begin to be approached.

**Table 1.1:** *The first few exact solutions for the hydrogen atom wavefunction.*

$n$	$l$		
	0		1
	$m = 0$	$m = 0$	$m = \pm 1$
1	$\left(\frac{1}{\pi a^3}\right)^{\frac{1}{2}} e^{(-\frac{r}{a})}$	—	—
2	$\frac{1}{\sqrt{8\pi a^3}} \left(1 - \frac{r}{2a}\right) e^{(-\frac{r}{2a})}$	$\frac{1}{\sqrt{32\pi a^3}} \frac{r}{a} e^{(-\frac{r}{2a})} \cos\theta$	$\mp \frac{1}{\sqrt{64\pi a^3}} \frac{r}{a} e^{(-\frac{r}{2a})} \sin\theta e^{\pm i\varphi}$

## 1.2 Hartree-Fock

While the wavefunction might appear complex, it is important to remember that it is still just a wave, and one of the fundamental properties of waves is that they can be added together to produce a separate, more complex wave. Likewise, any wave, no matter how complex, can be broken down as a summation of many simpler waves[5]. Thus, any linear combination of solutions to the wavefunction must therefore be a solution of a different (albeit perhaps non-physical) wavefunction. It is this property that will be exploited to solve our many electron problem. But first we must layout some ground rules.

The first rule is that the overall wavefunction must describe a system of fermions. This means that the wavefunction must be antisymmetric with respect to the exchanging of electronic coordinates, i.e.  $\Psi(1, 2, \dots, n, \dots, m) = -\Psi(1, 2, \dots, m, \dots, n)$  and the probability density of the wavefunction must be indistinguishable with respect to the exchanging of electronic coordinates, i.e.  $|\Psi(1, 2, \dots, n, \dots, m)|^2 = |\Psi(1, 2, \dots, m, \dots, n)|^2$ . The easiest way to meet this requirement is by using a Slater determinant[6], and by introducing the concept of spin. Spin is actually a fourth quantum number, which I have neglected to talk about so far. Spin is a confusing property that some quantum particles have. No one really knows what it is, other than to say it has something to do with the angular momentum of the particle, and oddly enough that it in no way refers to the direction that the particle is spinning. It has been shown to have binary values[7], which I will refer to as up spin ( $\uparrow$ ) and down spin ( $\downarrow$ ). A Slater determinant that obeys antisymmetry is shown in Equation 1.12

$$\Psi(1, 2, 3, \dots, n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(1) & \psi_2(1) & \psi_3(1) & \dots & \psi_n(1) \\ \psi_1(2) & \psi_2(2) & \psi_3(2) & \dots & \psi_n(2) \\ \psi_1(3) & \psi_2(3) & \psi_3(3) & \dots & \psi_n(3) \\ \dots & \dots & \dots & \dots & \dots \\ \psi_1(n) & \psi_2(n) & \psi_3(n) & \dots & \psi_n(n) \end{vmatrix} \quad (1.12)$$

Here,  $n$  refers to the number of electrons in the system, and the numbers in parentheses refer to specific electrons.  $\psi_i$  refers to the  $i^{th}$  spin orbital which is equal to  $\phi_{\frac{i+1}{2}}(r, \theta, \varphi)\alpha(\omega)$  if  $i$  is odd, or  $\phi_{\frac{i}{2}}(r, \theta, \varphi)\beta(\omega)$  if  $i$  is even ( $\phi$  is an orbital and  $\omega$  is the spin coordinate).  $\alpha$  and  $\beta$  are called spin functions. What they actually are doesn't matter, but they do have to be orthonormal to each other.

$$\begin{aligned} \langle \alpha(\omega_i) | \alpha(\omega_i) \rangle &= \langle \beta(\omega_i) | \beta(\omega_i) \rangle = 1 \\ \langle \alpha(\omega_i) | \beta(\omega_i) \rangle &= \langle \beta(\omega_i) | \alpha(\omega_i) \rangle = 0 \end{aligned} \quad (1.13)$$

Using Slater determinants has some interesting consequences. If we had a system where two electrons of the same spin were in the same spin orbital, that would mean that two of the columns would be equal to one another. Such a determinant would be equal to zero,

which means that the wavefunction for such a system would not exist. This is the famous Pauli antisymmetry principle[8]. The math showing that it satisfies antisymmetry is shown below[9].

$$\Psi(1, 2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1(1)\alpha(\omega_1) & \phi_1(1)\beta(\omega_1) \\ \phi_1(2)\alpha(\omega_2) & \phi_1(2)\beta(\omega_2) \end{vmatrix} \quad (1.14)$$

$$\Psi(1, 2) = \frac{1}{\sqrt{2}} [\phi_1(1)\alpha(\omega_1)\phi_1(2)\beta(\omega_2) - \phi_1(1)\beta(\omega_1)\phi_1(2)\alpha(\omega_2)] \quad (1.15)$$

$$-\Psi(1, 2) = \frac{1}{\sqrt{2}} [\phi_1(1)\beta(\omega_1)\phi_1(2)\alpha(\omega_2) - \phi_1(1)\alpha(\omega_1)\phi_1(2)\beta(\omega_2)] \quad (1.16)$$

$$-\Psi(1, 2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1(2)\alpha(\omega_2) & \phi_1(2)\beta(\omega_2) \\ \phi_1(1)\alpha(\omega_1) & \phi_1(1)\beta(\omega_1) \end{vmatrix} = \Psi(2, 1) \quad (1.17)$$

As can be seen, changing the indexing of the electrons is equivalent to swapping the rows for those electrons, which changes the sign of the determinant. It is important to note that because observable quantities depend not the wavefunction, but the *absolute square* of the wavefunction, altering the indexing of the electrons does not change the calculation of these observables. Consider the example of the electron density probability distribution shown below (note that  $\phi$ ,  $\alpha$ , and  $\beta$  are real in this example, but even if they weren't the results would not change).

$$|\Psi(1, 2)|^2 = \frac{1}{2} \left[ \begin{vmatrix} \phi_1(1)\alpha(\omega_1) & \phi_1(1)\beta(\omega_1) \\ \phi_1(2)\alpha(\omega_2) & \phi_1(2)\beta(\omega_2) \end{vmatrix}^2 \right] \quad (1.18)$$

$$\begin{aligned} |\Psi(1, 2)|^2 &= \frac{1}{2} [(\phi_1(1)\alpha(\omega_1)\phi_1(2)\beta(\omega_2) - \phi_1(1)\beta(\omega_1)\phi_1(2)\alpha(\omega_2)) \\ &\quad \times (\phi_1(1)\alpha(\omega_1)\phi_1(2)\beta(\omega_2) - \phi_1(1)\beta(\omega_1)\phi_1(2)\alpha(\omega_2))] \end{aligned} \quad (1.19)$$

$$\begin{aligned} |\Psi(1, 2)|^2 &= \frac{1}{2} \int \int \int \int [\phi_1(1)^2\alpha(\omega_1)^2\phi_1(2)^2\beta(\omega_2)^2 + \phi_1(1)^2\beta(\omega_1)^2\phi_1(2)^2\alpha(\omega_2)^2 \\ &\quad - 2\phi_1(1)^2\alpha(\omega_1)\beta(\omega_1)\phi_1(2)^2\beta(\omega_2)\alpha(\omega_2)] d1d2d\omega_1d\omega_2 \end{aligned} \quad (1.20)$$

$$|\Psi(1, 2)|^2 = \frac{1}{2} [|\phi_1(1)|^2|\phi_1(2)|^2 + |\phi_1(1)|^2|\phi_1(2)|^2] \quad (1.21)$$

### 1.2.1 Two-Electron Integrals

As stated before, the main difficulty of performing calculations on many electron systems is figuring out how to deal with the interactions between the electrons. We can get around this

by making the assumption that these interactions can be approximated by figuring out how a single electron interacts with the *average* field generated by all the other electrons. This is called the self-consistent-field (SCF) method[10]. With this, we can replace the Hamiltonian operator with the one-electron Fock operator which is given below[9]

$$\hat{f}(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} + v^{HF}(i) \quad (1.22)$$

The first two terms are combined into an operator called the  $H^{\text{core}}$  operator ( $\hat{h}(i)$ ) and  $v^{HF}(i)$  is the potential generated by all the electrons except the  $i^{\text{th}}$  one.  $v^{HF}(i)$  is composed of two different operators: the Coulomb operator and the exchange operator. The Coulomb operator is intuitive, it is simply the interaction electron 1 experiences when in a Coulomb potential generated by electron 2

$$\hat{J}_b(1)\psi_a(1) = \left[ \int \psi_b^*(2) \frac{1}{r_{12}} \psi_b(2) d\mathbf{x}_2 \right] \psi_a(1) \quad (1.23)$$

where  $J_b(1)$  is the Coulomb operator for electron 1 when interacting with an electron in spin orbital  $b$  and  $\mathbf{x}_i$  is the spatial coordinates of electron  $i$ . The exchange operator is slightly more complicated than the Coulomb operator. While the Coulomb operator can be defined without a function to operate on (simply remove  $\psi_a(1)$  from both sides of Equation 1.23 and it is still a valid operator), the exchange operator only is defined when operating on a function. Its form is given below.

$$\hat{K}_b(1)\psi_a(1) = \left[ \int \psi_b^*(2) \frac{1}{r_{12}} \psi_a(2) d\mathbf{x}_2 \right] \psi_b(1) \quad (1.24)$$

where  $K_b(1)$  is the exchange operator for electron 1 operating on spin orbital  $a$ . At first glance, it might appear to be almost the same as Equation 1.23, but notice that  $\psi_a$  has been swapped (or exchanged) with  $\psi_b$  *inside* the integral. The term itself does not really have a classical interpretation, but its need arises due to the use of an antisymmetric wavefunction.

Now it would be helpful to introduce some new notation. The expectation values for these operators are defined as follows

$$\langle \psi_a(1) | \hat{J}_b(1) | \psi_a(1) \rangle = \int \int \psi_a^*(1) \psi_a(1) \frac{1}{r_{12}} \psi_b^*(2) \psi_b(2) d\mathbf{x}_1 d\mathbf{x}_2 = (aa|bb) \quad (1.25)$$

$$\langle \psi_a(1) | \hat{K}_b(1) | \psi_a(1) \rangle = \int \int \psi_a^*(1) \psi_b(1) \frac{1}{r_{12}} \psi_b^*(2) \psi_a(2) d\mathbf{x}_1 d\mathbf{x}_2 = (ab|ba) \quad (1.26)$$

These are called the two-electron integrals. The Hartree-Fock equation can be written as

$$\left[ \hat{h}(1) + \sum_{b \neq a}^n (J_b(1) - K_b(1)) \right] \psi_a(1) = \epsilon_a \psi_a(1) \quad (1.27)$$

$$\langle \psi_a(1) | \hat{h}(1) | \psi_a(1) \rangle + \sum_{b \neq a}^n [(aa|bb) - (ab|ba)] = \epsilon_a \quad (1.28)$$

where  $n$  is the number of spin orbitals in the system and  $\epsilon_a$  is the energy of an electron in spin orbital  $a$ . If instead we are interested in the energy of an entire atom or molecule, the equation becomes

$$\sum_a^n \langle \psi_a(1) | \hat{h}(1) | \psi_a(1) \rangle + \frac{1}{2} \sum_a^n \sum_b^n [(aa|bb) - (ab|ba)] = \epsilon \quad (1.29)$$

where the factor of one half is multiplied to the two electron terms as a result of the interactions being double counted by the summations.

## 1.2.2 Basis Sets

Back in section 1.2, I said that there were some ground rules for our wavefunction. The second of these ground rules is that our trial wavefunction ( $\Phi$ ) must meet the boundary conditions for the original problem. Up until now, I have used only spin orbitals ( $\psi$ ) to describe the wavefunction, which as you'll recall is composed of a spin function, which I have already talk about at length, and a spatial orbital ( $\phi$ ) which I have almost completely ignored. I'll discuss the spatial orbitals next.

As I mentioned, the overall wavefunction of a system ( $\Psi$ ) cannot be known exactly unless we are looking only at hydrogen-like atoms. So instead we write the wavefunction as a Slater determinant composed of spin orbitals. Okay, so what exactly is the mathematical form of these spin orbitals? Well, as it turns out, unless we are talking only about atoms, we have no idea what these are either, even from a numerical stand point. We can make a fairly educated guess about what they *should* resemble though. The general idea is this: by pre-calculating a set of functions with adjustable parameters ( $\zeta$ ), we can produce a wavefunction with a linear combination of  $\Phi(\zeta)$  that adequately reproduces results of numerical calculations of  $\phi$ . Mathematically, this is expressed as[9]

$$\phi_u \approx \sum_{i=1}^K C_{u,i} \Phi(\zeta_{u,i}) \quad (1.30)$$

where  $\zeta_{u,i}$  is the  $i^{th}$  adjustable parameters of the basis function  $\Phi$  which is approximating the  $u^{th}$  orbital,  $C_{u,i}$  is the  $i^{th}$  coefficient of  $i^{th}$  basis function of the  $u^{th}$   $\phi$ , and  $K$  is the total number of basis functions that make up the basis set.  $C_{u,i}$  can intuitively be thought of as



the “percentage” of  $\phi$  that  $\Phi(\zeta_{u,i})$  makes up (this is not strictly speaking true, as  $C_{u,i}$  can be greater than one or negative, but the general idea is there).

So how do we choose our  $\Phi$ s? This is where our ground rule comes in to play. We must choose a function that meets the boundary conditions of the problem, which means that it must go to zero at infinity (this is also not exactly true, we can choose any function we want so this rule is really more of a guideline, but our basis sets will be much smaller the closer they are to the actual problem so it is good practice to follow it). Because an atom (or molecule) consists of electrons in the potential generated by one (or many) nuclei, it is reasonable to conclude that the wavefunction would be similar to that of the hydrogen atom. We might initially think to use the exact hydrogen atom solution, but it quickly requires lots of floating point operations to compute which makes it impractical to use for large scale calculations. A more reasonable function is the Slater-type orbital (STO)[11]. The radial part of its 1s orbital centered at  $R_A$  is shown below

$$\Phi_{1s}^{STO}(r - R_A) = \sqrt{\frac{\zeta^3}{\pi}} e^{-\zeta|r-R_A|} \quad (1.31)$$

This is more well behaved than the exact hydrogen atom solution, but problems arise when we try to use them in the two-electron integrals. Because we may be trying to do a double integral over function which could have up to four different centers, we end up with something that is very hard to compute. Therefore, Gaussian-type orbitals (GTOs) have become the de facto method of producing basis sets. The radial part of its 1s orbital centered at  $R_A$  is shown below[12]

$$\Phi_{1s}^{GTO}(r - R_A) = \left(\frac{2\zeta}{\pi}\right)^{\frac{3}{4}} e^{-\zeta|r-R_A|^2} \quad (1.32)$$

Because the product of two different Gaussians centered at two different locations is a third Gaussian centered at a third location, this makes the two-electron integrals much easier to deal with. The trade off is that GTOs do not as accurately match the hydrogen atom solution as STOs, but this can be fixed somewhat by just using more of them.

All of this reduces our wavefunction issue to a simple optimization problem, which will be discussed more in-depth in Chapter 3.

### 1.2.3 Hartree-Fock SCF

In the previous sections, we discussed the calculation of the two-electron integrals and introduced the concept of using basis sets to approximate a wavefunction. Now we will discuss how they are actually used.

SCF begins with the formation of the density matrices. If we are looking at a single atom, we can exploit the symmetry of the problem by having a smaller density matrix for each orbital symmetry ( $\lambda$ ), instead of one large one for the whole system. There are also two types of density matrices, one for contributions from closed shell spinors ( $\mathbf{D}_C$ ) and one for contributions from open shelled spinors ( $\mathbf{D}_O$ ).

Using a procedure developed by Roothaan and Hall[10, 13], we first need to find the matrix  $\mathbf{C}$  which satisfies

$$\mathbf{F}\mathbf{C} = \epsilon\mathbf{S}\mathbf{C} \quad (1.33)$$

where  $\mathbf{F}$  is the Fock matrix,  $\epsilon$  is the vector of eigenvalues of  $\mathbf{F}$  with respect to matrix  $\mathbf{S}$ , and  $\mathbf{S}$  is the overlap matrix. The matrix elements of  $\mathbf{S}$  are

$$\mathbf{S}_{p_\lambda, q_\lambda} = \langle p_\lambda | q_\lambda \rangle \quad (1.34)$$

and the non-relativistic  $\mathbf{F}$  for closed shells of symmetry  $\lambda$  is given by

$$\mathbf{F}_{p_\lambda, q_\lambda} = \langle p_\lambda | \hat{h}(1) | q_\lambda \rangle + \sum_{\mu=1}^{nsym} \sum_{r=1}^K \sum_{s=1}^K \left( 2 \sum_{a=1}^{occ} \mathbf{C}_{r_\mu, a} \mathbf{C}_{s_\mu, a}^* \right) \left[ (p_\lambda q_\lambda | r_\mu s_\mu) - \frac{1}{2} (p_\lambda q_\lambda | s_\lambda r_\lambda) \right] \quad (1.35)$$

where  $p$ ,  $q$ ,  $r$ , and  $s$  are basis functions,  $K$  is the total number of basis functions for symmetry  $\mu$ , and  $occ$  is the number of occupied closed shells in symmetry  $\mu$ . Because  $p$  and  $q$  will always be of the same symmetry, and likewise with  $r$  and  $s$ , the  $\lambda$  and  $\mu$  subscripts will be dropped from these indices and instead will be obtained from context. As we can see,  $\mathbf{F}$  is itself a function of  $\mathbf{C}$  and we left with a rather disappointing outcome. That is, in order for us to find  $\mathbf{C}$ , we first must know what it is. And so it would seem that all is lost and all of this hard work was for nothing. And yet, there is a glimmer of hope.  $\mathbf{F}$  is made up of two terms, and as luck would have it, only the term corresponding to contributions from the two-electron integrals depends on  $\mathbf{C}$ , the term from one-electron integrals can be used independently of  $\mathbf{C}$ . Thus, we begin by assuming that  $\mathbf{F}$  depends solely on the one-electron term, effectively assuming that  $\mathbf{C} = \mathbf{0}$ . Solving Equation 1.33 with the  $H^{\text{core}}$  hamiltonian matrix only will give us the initial guess of  $\mathbf{C}$ . With this, we can then resolve  $\mathbf{F}$  with the guess of  $\mathbf{C}$  and diagonalize the new  $\mathbf{F}$  matrix to get another set of values for  $\mathbf{C}$ . We can then repeat this process over and over until  $\mathbf{C}$  is no longer changing outside an acceptable tolerance at which point we say the calculation has converged. We have now just completed a Hartree-Fock calculation.

## 1.3 Special Relativity

Another field of physics that was being developed at the same time as quantum mechanics was special relativity (general relativity which involves gravity's effect on spacetime is far, *far* outside the scope of this thesis. I will therefore be refining to special relativity as just relativity from now on). Relativity traces its roots to Michelson and Morley[14] when they made the startling observation that all observers, no matter their frame of reference, will always agree on the speed of light. But it would not be until everyone's favorite Swiss patent cleric decided to examine this issue that the truly profound implications of this observation would be known.

This is perhaps a bit of an aside, but it is simply too interesting to pass up. Consider this thought experiment: you and I are in two different spaceships in space. We will define your frame of reference such that you are stationary at the origin, and that I am flying past you at a constant speed that is arbitrarily close to the speed of light, we will call this speed  $v$ . In my spaceship, there is a laser that I have pointed perfectly perpendicular to the direction I am traveling. From my frame of reference, where I am stationary, I turn the laser on and observe photons that move from the laser to the side of my spaceship, a distance equal to  $d_1$ , in time equal to  $t$  at the speed of light  $c$ . Back in your frame of reference, you are also watching my experiment, but you make some different observations. In the time it take for the photons to move across my spaceship, I move a small distance  $d_2$ . So the total distance you see the photons move is not  $d_1$ , but  $d_3$  which is the hypotenuse of the right angle triangle formed by  $d_1$ ,  $d_2$ , and  $d_3$ . *But*, we both agree that the photons were moving at speed  $c$ ! So we can not agree on the time it took to do so because  $d_3 > d_1$ . So you will observe the time it takes for the photons to move to be a time equal to  $t'$ . We can derive the relationship between  $t$  and  $t'$  in the following way.

$$d_1 = ct \qquad d_2 = vt' \qquad d_3 = ct' \qquad (1.36)$$

$$d_3^2 = d_1^2 + d_2^2 \qquad (1.37)$$

$$(ct')^2 = (vt')^2 + (ct)^2 \qquad (1.38)$$

$$(ct)^2 = (ct')^2 - (vt')^2 \qquad (1.39)$$

$$t = t' \sqrt{1 - \frac{v^2}{c^2}} \qquad (1.40)$$

Not does this give the absolutely remarkable result that time is relative, Equation 1.40 also implies that an object can have a maximum speed. But how could this be possible, if I turned on the engine on my spaceship, what's to stop me from just getting faster and faster? The answer is that the mass of an object increases the closer to the speed of light it is. This

can be described by[15]

$$m(v) = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (1.41)$$

where  $m(v)$  is the mass of an object moving at speed  $v$  and  $m_0$  is the rest mass of the object. With this we can see how it would be impossible for an object with mass to move at the speed of light. Its mass would go to infinity and would therefore need an infinite amount of energy to get there.

If we expand Equation 1.41 as a Taylor series centered at zero, we obtain the following

$$m(v) = m_0 \left[ 1 + \frac{1}{2} \frac{v^2}{c^2} + \frac{3}{8} \frac{v^4}{c^4} + \frac{5}{16} \frac{v^6}{c^6} + \frac{35}{128} \frac{v^8}{c^8} + \dots \right] \quad (1.42)$$

If we rearrange this, and multiply both sides by  $c^2$  we get

$$(m(v) - m_0)c^2 = \frac{m_0 v^2}{2} + m_0 \left[ \frac{3}{8} \frac{v^4}{c^2} + \frac{5}{16} \frac{v^6}{c^4} + \frac{35}{128} \frac{v^8}{c^6} + \dots \right] \quad (1.43)$$

So if we take the mass of an object at speed  $v$  and subtract the mass of the object and then multiply the result by  $c^2$ , we get the kinetic energy of the object (plus some terms that very quickly go to zero). This is what led Einstein to believe that the total kinetic energy of an object was stored in its mass. He would use this to produce the world's most famous equation

$$E_{kin} = m(v)c^2 \quad (1.44)$$

### 1.3.1 Relativistic Quantum Mechanics

So it would seem that relativity concerns the physics of the very large and fast, while quantum mechanics is concerned with the very small. How do we merge the two. So lets start from the beginning with the time-dependent one-dimensional Schrödinger equation which I will rewrite here in a more general and explicit form[1]

$$i\hbar \frac{\partial \Psi(x, t)}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + \hat{V}(x)\Psi(x, t) \quad (1.45)$$

where  $i\hbar \frac{\partial}{\partial t}$  is the energy operator,  $-\hbar^2 \frac{\partial^2}{\partial x^2}$  is the momentum squared operator ( $\hat{p} = -i\hbar \frac{\partial}{\partial x}$ ),  $\hbar$  is the reduced Plank's constant ( $\hbar = \frac{h}{2\pi}$ ) and  $\hat{V}(x)$  is the potential energy operator that the particle described by  $\Psi(x, t)$  experiences. The reason we can not use the Schrödinger equation directly in relativistic quantum mechanics is because it treats time

and space differently (it depends on only the first derivative of time, but the second derivative of space). So we need some way to fix this. If we try to express Equation 1.44 in terms of  $\hat{p}$  we get[15]

$$E_{kin}^2 = \hat{p}^2 c^2 + m_0^2 c^4 \quad (1.46)$$

If we want to solve for a free particle ( $\hat{V}(x) = 0$ ), we could put the energy and momentum operator from above into Equation 1.46 and multiply both sides from the right to obtain

$$-\hbar^2 \frac{\partial^2 \Psi(x, t)}{\partial t^2} = c^2 \left( -\hbar^2 \frac{\partial^2 \Psi(x, t)}{\partial x^2} + m_0^2 c^2 \Psi(x, t) \right) \quad (1.47)$$

which is called the Klein-Gordon[16, 17] equation for a free particle. We could also solve for a particle of charge  $q$  in a electromagnetic field described by potentials  $\hat{A}$  and  $\hat{\varphi}$  as

$$\left( i\hbar \frac{\partial}{\partial t} - q\hat{\varphi} \right)^2 \Psi(x, t) = c^2 \left[ \left( -i\hbar \frac{\partial}{\partial x} - \frac{q}{c} \hat{A} \right)^2 + m_0^2 c^2 \right] \Psi(x, t) \quad (1.48)$$

Here we are treating time and space exactly the same, but there is a problem. While this equation works perfectly well for spinless particles, we can not ad hoc add spin to this equation like we did with the original Schrödinger equation without violating its Lorentz invariance, which is also important for relativistic theories. Also, because this equation solves for the *square* of the energy, it implies that the energy could both be positive or negative. This is a puzzling result that we will return to later, but for now let's examine how Dirac fixed the issue of introducing spin.

### 1.3.2 Dirac Equation

Dirac[18] began with the Klein-Gordon equation which I will rewrite in its three-dimensional form as

$$\pi_0^2 - \left[ \sum_{u=x,y,z} \pi_u^2 + m_0^2 c^2 \right] = 0 \quad (1.49)$$

$$\pi_0 = \frac{i\hbar \frac{\partial}{\partial t} - q\hat{\varphi}}{c}, \quad \pi_u = -i\hbar \frac{\partial}{\partial u} - \frac{q}{c} \hat{A} \quad (1.50)$$

Dirac wanted to write this equation as the product of one equation and its conjugate, which can be done as follows

$$\left( \pi_0 + \left[ \sum_{u=x,y,z} \alpha_u \pi_u + \alpha_0 m_0 c \right] \right) \left( \pi_0 - \left[ \sum_{u=x,y,z} \alpha_u \pi_u - \alpha_0 m_0 c \right] \right) = 0 \quad (1.51)$$

so long as we restrict  $\alpha$  as

$$\alpha_i^2 = 1, \quad \alpha_i \alpha_j + \alpha_j \alpha_i = 0 \quad \text{for } i \neq j \quad (1.52)$$

It's clear that the  $\alpha$ s must be matrices. The simplest of these that meet the requirements of Equation 1.52 are

$$\alpha_i = \begin{pmatrix} \mathbf{0} & \sigma_i \\ \sigma_i & \mathbf{0} \end{pmatrix} \quad \text{for } i \neq 0 \quad (1.53)$$

$$\alpha_0 \equiv \beta = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix} \quad (1.54)$$

where  $\sigma_i$  are the Pauli spin matrices for  $i = x, y, z$ ,  $\mathbf{0}$  is a  $2 \times 2$  matrix of zeros, and  $\mathbf{I}$  is a  $2 \times 2$  identity matrix[15]. We now have two different operator equations

$$\left( \pi_0 + \sum_{u=x,y,z} \alpha_u \pi_u + \beta m_0 c \right) \Psi(x, y, z, t) = 0 \quad (1.55)$$

$$\left( \pi_0 - \sum_{u=x,y,z} \alpha_u \pi_u - \beta m_0 c \right) \Psi(x, y, z, t) = 0 \quad (1.56)$$

Equation 1.55 will give negative energy and Equation 1.56 will give positive energy. The positive energy solutions are electrons, but what about the negative energy solutions. As it turns out, they correspond to antimatter, which is remarkable because antimatter would not be discovered until years later[19]. The stationary states (time-independent) for the electron then become

$$\left( E - V - \beta m_0 c^2 - c \sum_{u=x,y,z} \alpha_u \pi_u \right) \Psi(x, y, z) = 0 \quad (1.57)$$

where  $V = q\hat{\phi}$ . Introducing  $\alpha$  as a  $4 \times 4$  matrix means that  $\Psi$  must also have four components. It is typically written as

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\psi} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\phi} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix} \quad (1.58)$$

where  $\boldsymbol{\psi}$  and  $\boldsymbol{\phi}$  are called spinors.  $\boldsymbol{\psi}$  is called the large component and  $\boldsymbol{\phi}$  is the small component. With this in hand, we are almost done, but there is one more trick we must use in-order for the Dirac equation to be usable.

### 1.3.3 Kinetic Balancing

Recall that Equation 1.57 is only half of the total equation. We still have all of the negative energy solutions to deal with. If we were to try and use Equation 1.57 as is, the numerical methods we use of solving it would find these negative energies, and as there is no lower bound on the negative solutions, our solutions to the equation would all drop to negative infinity. We can solve this problem if we assume that all the infinite amount of negative energy states are filled with an infinite amount of positrons. This means that our electron can not drop into the negative energies due to the Pauli exclusion principal. Mathematically, we can reproduce this idea using a procedure called kinetic balancing, which is done in the following way[15].

Using the spinors, we can write Equation 1.57 as a matrix multiplication problem of the form

$$\begin{pmatrix} (V + \beta m_0 c^2) & c(\boldsymbol{\sigma}\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\boldsymbol{\pi}) & (V + \beta m_0 c^2) \end{pmatrix} \begin{pmatrix} \psi \\ \phi \end{pmatrix} = \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix} \begin{pmatrix} \psi \\ \phi \end{pmatrix} \quad (1.59)$$

where  $\boldsymbol{\sigma}\boldsymbol{\pi} = \sum_{u=x,y,z} \alpha_u \boldsymbol{\pi}_u$ . This is equivalent to adding the following two equations together

$$(E - V - m_0 c^2)\psi - c(\boldsymbol{\sigma}\boldsymbol{\pi})\phi = 0 \quad (1.60)$$

$$(E - V + m_0 c^2)\phi - c(\boldsymbol{\sigma}\boldsymbol{\pi})\psi = 0 \quad (1.61)$$

We can set where the energy is zero to anywhere we want, so lets make zero the energy of a free electron at rest ( $\epsilon = E - m_0 c^2$ ). Equations 1.60 and 1.61 now become

$$(\epsilon - V)\psi - c(\boldsymbol{\sigma}\boldsymbol{\pi})\phi = 0 \quad (1.62)$$

$$(\epsilon - V + 2m_0 c^2)\phi - c(\boldsymbol{\sigma}\boldsymbol{\pi})\psi = 0 \quad (1.63)$$

We can rewrite Equation 1.63 as

$$\phi = \left(1 + \frac{(\epsilon - V)}{2m_0 c^2}\right)^{-1} \frac{1}{2m_0 c^2} \boldsymbol{\sigma}\boldsymbol{\pi}\psi \quad (1.64)$$

To a good approximation,  $\epsilon - V$  will be much smaller than  $2m_0 c^2$ , so we can assume the term in parentheses is equal to 1. Therefore, the small component can be written in terms of just  $\psi$  and becomes

$$\phi = \frac{\boldsymbol{\sigma}\boldsymbol{\pi}}{2m_0 c^2} \psi \quad (1.65)$$

Not only does this make solving the Dirac equation much simpler (we only need to solve for  $\psi$  instead of  $\psi$  and  $\phi$ ) it also “hides” the negative energies from numerical methods.

Note, though, that it relies on the assumption that  $\epsilon - V$  is much smaller than  $2m_0c^2$ . If this is not the case, such as in the electrons in the heaviest of known elements, then this procedure will fail and we will get energies that are lower than they should be. When this occurs, we say that there is variational prolapse of the wavefunction.

### 1.3.4 The Dirac Hydrogen-Like Atom

Up until now, we have only looked at what Dirac's theory says about a free electron (well, a "free" electron in the presence of an infinite amount of positrons, but I digress). So what happens to the hydrogen-like atom solutions?. The quantum numbers we get here are similar to the non-relativistic theory. They are: the principal quantum number ( $n = 1, 2, 3, \dots$ ), the azimuthal quantum number ( $l = 0, 1, 2, \dots, n-1$ ), the angular momentum quantum number ( $j = |l \pm \frac{1}{2}|$ ), and the magnetic quantum number ( $m = -j, -j+1, \dots, j-1, j$ ). The energy in terms of the quantum numbers is given by[15]

$$E_{n,j} = -\frac{1}{2n^2} \left[ 1 + \frac{1}{nc^2} \left( \frac{1}{j + \frac{1}{2}} - \frac{3}{4n} \right) \right] \quad (1.66)$$

It can be seen that there is a splitting in the energy levels of spinors with the same azimuthal number, whereas in the non-relativistic theory, orbitals with the same azimuthal number would all be perfectly degenerate.

### 1.3.5 Relativistic Effects

The effects of relativity have a dramatic effect on chemistry. Two classic examples are gold's distinctive colour, and mercury being the only metal that is liquid at room temperature. In the case of gold, its  $6s_{1/2}$  electrons are lowered in energy and its  $5d_{5/2}$  electrons are raised. This shifts its absorption region into the blue end of the spectrum, making it appear yellow[20]. The  $6s_{1/2}$  electrons of mercury are similarly stabilized and since its  $5d_{3/2}$  and  $5d_{5/2}$  orbitals are filled, it behaves quite like a noble gas, making bonding quite difficult at room temperature[21].

Relativistic effects are broken up into two categories: primary and secondary effects. Primary effects are due to the contraction of the s and p electrons. The non-relativistic Bohr radius for the 1s electron of hydrogen is[3]

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_0e^2} \quad (1.67)$$

where  $\epsilon_0$  is the permittivity of free space,  $\hbar$  is the reduced planks constant,  $e$  is the charge of an electron, and  $m_0$  is the mass of an election at rest. Recall from Equation 1.41 that the



mass of an object increases with its speed. If we substitute  $m(v)$  from Equation 1.41 for  $m_0$  into Equation 1.67 we get

$$a_v = \frac{4\pi\epsilon_0\hbar^2\sqrt{1 - \frac{v^2}{c^2}}}{m_0e^2} \quad (1.68)$$

Dividing Equation 1.68 by 1.67 gives

$$\frac{a_v}{a_0} = \sqrt{1 - \frac{v^2}{c^2}} \quad (1.69)$$

which is less than 1 and will get closer and closer to zero as  $v$  increases ( $v$  will increase in proportion to the nuclear charge). This contraction stabilizes the electrons, making them more inert.

Secondary effects are a direct result of the electron contraction. The contraction of these electrons leads to greater shielding of the nucleus, reducing its effective nuclear charge. This causes the d and f electrons to feel less attracted to their nucleus, which makes their orbitals enlarge. This destabilizes these electrons, making them more reactive[15].

### 1.3.6 Gaussian Nucleus

Up till now, we have assumed the nucleus can be represented as a point. For many applications, this works just fine, but it can make it difficult to find convergence when relativity is considered. Therefore, it is helpful to use a model more grounded in reality in this case. One such model is the Gaussian nucleus. This changes the potential in Equation 1.59 to

$$V(\mathbf{r}) = Z \left( \frac{\alpha}{\pi} \right)^{\frac{3}{2}} \int \frac{e^{-\alpha(\mathbf{r}-\mathbf{R})^2}}{|\mathbf{r}-\mathbf{R}|} d\mathbf{R} \quad (1.70)$$

where  $\mathbf{r}$  is the coordinates of an electron,  $\mathbf{R}$  is the coordinates of the nuclei,  $Z$  is the nuclear charge, and  $\alpha$  is an estimation of the “size” of the nucleus. This is usually approximated using the root-mean-square (RMS) radius of the nuclei[22], which can be related to the mass of the nucleus by[23]

$$\sqrt{\langle R^2 \rangle} = (0.836A^{1/3} + 0.570) \quad (1.71)$$

where  $\langle R^2 \rangle$  is the mean-square radius in femtometers, and  $A$  is the mass of the nucleus. Converting the RMS into atomic units and substituting it into the following equation gives the value of  $\alpha$ [22]

$$\alpha = \frac{3}{2\langle R^2 \rangle} \quad (1.72)$$

# Chapter 2

## cudaDFRATOM

### 2.1 GPU Architecture

While general purpose computing on graphics processing units (GPGPU) (GPUs) has been adopted by the high performance computing (HPC) community for quite some time, it can seem quite complex to the uninitiated. While most quantum chemists who decide to dip their toes into computational waters can get away with having little to no understanding of what is actually going on under the hood when programming for a central processing unit (CPU), the same can not at all be said for GPUs. Therefore, in this section I will explain what makes a GPU tick in order to help understand the terms and techniques used in the following chapter. As cudaDFRATOM was optimized for the Nvidia Tesla C2050, I will be referring to the Tesla C2050 specifications for examples when needed. I will begin with the most granularity possible, and zoom out so that by the end of this section, the reader should come away prepared for the rest of this chapter. The bulk of this section has been adapted from [24, 25] and from my own experience.

#### 2.1.1 Threads, Blocks, and Grids

The most granular element of computation on a GPU is the thread. When a CUDA function or subroutine (hereafter called a kernel) is called, it is a thread that actually executes the code. What makes GPGPU so powerful is that while a thread in a kernel is always executing the same code, the data a thread works with can be completely different between threads. This method of parallelism is called single instruction, multiple data (SIMD).

32 threads are organized into a structure called a warp. Within a warp, all threads execute code in lock-step. If a condition arises where some threads in warp must execute some code, and other threads in the same warp execute some other code (for instance in a **if else** statement), this leads to serialization of code execution, a process called warp divergence. In a worst case scenario, this could cause all 32 threads to execute serially which

could lead to a massive hit to performance. It can also cause warps to fall behind other warps and has the potential to cause race conditions to appear. If one warp needs data generated by another warp that is several steps behind, this can cause all kinds of confusing errors to appear. Therefore, it is generally advisable to avoid warp divergence whenever possible. But, for programs of any considerable usefulness, warp divergence will be inevitable. Thankfully, the good people of Nvidia have included the **syncthreads** command. This provides a barrier that all threads within a block must meet before any can continue.

The next highest structure is the thread block, or block for short. Simply put, a block is a collection of one or more threads. When a kernel is called, the block scheduler assigns each block to a streaming multiprocessor (SM). How many blocks a SM can run at once depends on the resources each block uses (more on this in section 2.1.2), but no more than 8 blocks per SM can ever be run at once on our C2050. Blocks also add an element of dimensionality to threads through the thread index (*threadIdx*) variable. This variable has three parts, *threadIdx.x*, *threadIdx.y*, and *threadIdx.z*. These variables make dealing with matrixes and arrays much simpler and can help with paralyzing large nested loops. The size of dimensions of the grid can be accessed through the *blockDim* variable. *blockDim* also has *x*, *y*, and *z* parts. These variables are all integer type, and the *threadIdx* variables range from 1 in Fortran (or 0 in C/++) to the relevant part of *blockDim* in Fortran (or *blockDim* -1 in C/++). The *threadIdx* set of variables are all reset at the boundary of each block.

Finally there is the grid. The grid is the entire collection of blocks that are launched for a kernel. They also give a dimensionality to blocks through the *blockIdx* variable which has similar parts to the *threadIdx* variable. It also adds in the *gridDim* variable which is analogous to *blockDim*.

## 2.1.2 Memory

To simplify the discussion of the kinds of memory available on a GPU we will limit ourselves to the three main types. They are the global memory, the shared memory, and the register memory. Because the execution time of a calculation is often not limited by the actual number of floating point operations to perform, but rather the reading and writing to memory, it is essential to be sure that the three main types of memory are managed correctly.

Global memory is both the most plentiful and the slowest available memory on a GPU. In a typical calculation, global memory will be allocated in code executed by the CPU (host). Data will be uploaded to the allocated memory and then used by the GPU (device). Global memory is distinct from the other kinds of device memory in that it can be read from and wrote to by all threads currently executing. This can allow threads of one block to pass messages to threads in an other block, but great care must be taken to avoid race conditions.

There is an additional caveat with global memory in that if threads are reading from non-contiguous memory locations, then memory bandwidth plummets. There is also a problem when multiple threads try to read from the same location. Instead of being one transaction, the read is serialized which leads to a massive performance penalty. It is therefore to the programmers advantage to write their kernels with as little use of the global memory as possible.

Shared memory, as its name implies, is shared and private to threads in a single block. It is able to be access much faster than global memory, but this comes at the cost of a much reduced volume. The amount of it available per SM varies with hardware generation, but the maximum available on our card is 48KB. This amount is shared among all blocks that are running on the SM, so the amount available per block will typically be even less. Also, there is no guarantee that memory loaded by one thread will be available to all other threads right away, only a call to **syncthreads** ensures proper loading. But, when used properly, shared memory is indispensable for speeding up calculations. It does not suffer from threads reading from non-contiguous memory locations like global memory does, and if all threads in a warp read from the same memory location a “broadcast” occurs. This means that instead of the read being serialized, all threads are able to read the data in one transaction. Shared memory is also distinct from global memory in that instead of being declared in the host code, it is declared in the device code. This is done by giving the variable the “shared” attribute in its declaration. One must be careful not to over use the shared memory however. As mentioned, there is only 48KB available per SM. If a block needs more than is available, the function will fail to execute. It can also affect the occupancy of the kernel, a concept that will be explained later on.

Lastly there is the register memory. This is the fastest memory available and is also private to an individual thread. As this may have been a lot to take in, Figure 2.1 shows the hierarchy of the grid, blocks, and threads, as well as how each memory type is localized. On the Tesla C2050, there are a total of 32768 32-bit registers available per SM. An individual thread in a block can use no more than 63 registers, and if the number of threads in a block times the number of registers needed per thread for a kernel is larger than the total registers available per SM then the execution of the kernel will fail. If a thread needs more memory than the registers will allow, then “register-spilling” takes place. This means that instead of a variable being stored into a register, it “spills over” into a fourth memory type called local memory. Local memory is really just a section of the global memory and comes with all the problems that global memory has, but the compiler is able to arrange it such that it can be read efficiently without the programmer having to worry about this. There is no way for a programmer to specify which variables will be stored in register or local memory, the compiler handles all of this. There are compiler flags which can change the maximum

amount of registers per thread allowed which will prompt the compiler to place more data into local memory. While this may sound like a bad idea, we will see in the next section how there are situations where this might be useful.

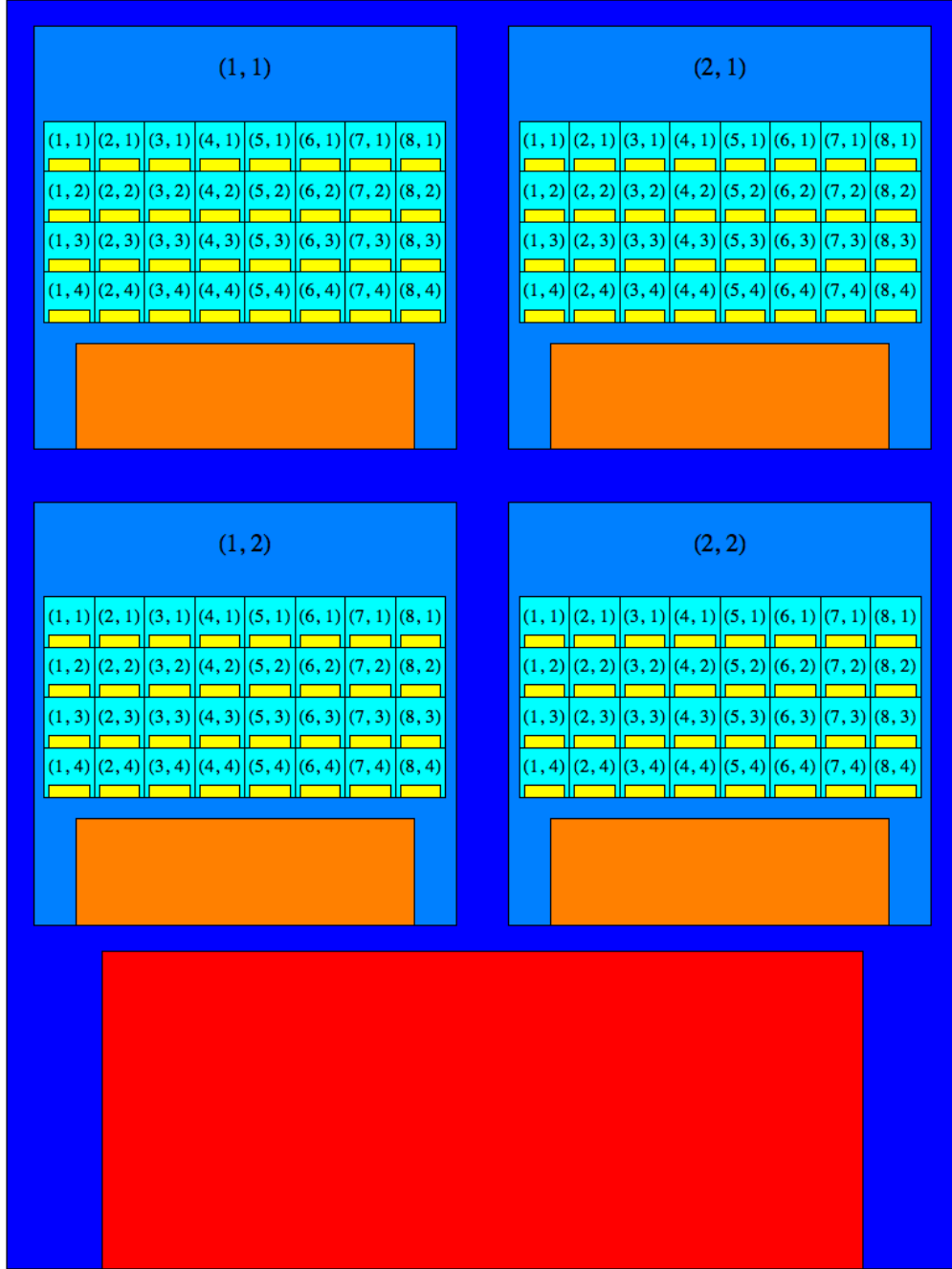
### 2.1.3 Occupancy and Performance

With the description of the various resources out of the way, now we can talk about how they can be used most effectively. The occupancy of a kernel is the number of warps that are actually able to run on a SM divided by the maximum possible warps a SM can run. Occupancy is important, as it determines the maximum number of blocks, and thus threads that are able to be run simultaneously. The occupancy of a kernel is determined by three things: the block size, the amount of shared memory per block, and the number of registers used per thread. While block size and shared memory are explicitly set by the programmer, it is next to impossible to guess how registers will be used before compile time. Thankfully, there are compiler flags that will display this information.

Armed with the resources our kernel will need, we can now determine its occupancy. The “CUDA Occupancy Calculator” is an Excel spreadsheet provided by NVIDIA and is an indispensable tool. We can input the needed resources and the hardware generation of our card and then receive the occupancy of the current resource configuration, where the bottle neck of the current configuration is, as well as information on how the occupancy will change upon tweaking the configuration. In some sense, the best block size is determined by how the other resources are used. Therefore, in order to maximize the occupancy, we must first make sure that we are using the memory provided correctly.

Registers are the fastest memory available, but they are also the easiest to over use. Therefore, they are best suited for small variables that need to be referenced frequently, such as accumulators. Shared memory is best used for storing large sections of non-contiguous data stored in the global memory. Finally data that is able to be stored nicely, or is infrequently used might best be left in the global memory.

As with everything, these rules are only guide lines. The best resource configuration will be individual not only to every kernel, but the GPU it is to be run on as well. The best way to maximize the efficiency of a calculation is to make minor alterations to where data is stored and play around with the code. Finally, it might not always be the best idea to maximize the occupancy of a kernel. If the bottleneck is not the actual calculation, but how fast the data can be read, then it might be advantageous to fill up the registers and shared memory as much as possible and have a very low occupation. The only way to know is to try!



**Figure 2.1:** Architecture of a GPU. This figure shows how threads and blocks are organized in the grid, as well as how the different memory types are localized within them. The grid is coloured dark blue, blocks are coloured light blue, and threads are coloured cyan. As for memory, the global memory is coloured red, shared memory is coloured orange, and register memory is coloured yellow. The numbers in parenthesis indicate the block or thread's `blockIdx` or `threadIdx` variables. In this particular case,  $\text{blockDim} = (8, 4)$  and  $\text{gridDim} = (2, 2)$ . A thread is able to access memory that is inside its registers, but not the registers of other threads, and memory that is inside the shared memory of its block, but not the shared memory of other blocks. All threads can access all of the global memory.

## 2.1.4 Matrix Multiplication Example

To complete the introduction to CUDA programming, we will go over a real example of CUDA Fortran code. For our example, we will examine code that performs matrix multiplication on matrices of arbitrary dimensions. This makes use of all the concepts we've covered so far, so it will be a good exercise to examine. The code is given below.

```
1 module mat_mul
2   implicit none
3   integer, parameter :: dp = kind(0.d0)
4   integer, parameter :: block_dim = 16
5   integer, constant :: rowA, colA, rowB, colB
6   contains
7     attributes(global) subroutine mat_mul_kernel(inputA, inputB, output)
8       real(dp), dimension(:, :) :: inputA, inputB, output
9       real(dp) :: total
10      real(dp), shared, dimension(block_dim, block_dim) :: s_tileA, s_tileB
11      integer :: x, y, i, j
12
13      x = threadIdx%x + blockDim%x * (blockIdx%x - 1)
14      y = threadIdx%y + blockDim%y * (blockIdx%y - 1)
15      total = 0.0_dp
16
17      do i = 0, colA - 1, block_dim
18        call syncthreads()
19        if ((x <= rowA) .and. (threadIdx%y + i <= colA)) then
20          s_tileA(threadIdx%x, threadIdx%y) = inputA(x, threadIdx%y + i)
21        else
22          s_tileA(threadIdx%x, threadIdx%y) = 0.0_dp
23        endif
24
25        if ((threadIdx%x + i <= rowB) .and. (y <= colB)) then
26          s_tileB(threadIdx%x, threadIdx%y) = inputB(threadIdx%x + i, y)
27        else
28          s_tileB(threadIdx%x, threadIdx%y) = 0.0_dp
29        endif
30
31        call syncthreads()
32
33        do j = 1, block_dim
34          total = total + s_tileA(threadIdx%x, j) * s_tileB(j, threadIdx%y)
35        enddo
36
37      enddo
38
39      if((x <= rowA) .and. (y <= colB)) then
40        output(x, y) = total
```

```

41         endif
42     end subroutine mat_mul_kernel
43 end module mat_mul
44
45 program main
46     use cudafor
47     use mat_mul
48     implicit none
49     ! Host variables
50     real(dp), allocatable, dimension(:, :) :: inputA, inputB, output
51     ! Device variables
52     real(dp), device, allocatable, dimension(:, :) :: d_inputA, d_inputB,
        d_output
53     ! For generating random numbers
54     integer, allocatable, dimension(:) :: seed
55     integer :: n
56     ! block and grid dimensions
57     type(dim3) :: blockDim, gridDim
58
59     rowA = 1024
60     colA = 432
61     rowB = colA
62     colB = 1783
63
64     ! Set up random number generator
65     call random_seed(size = n)
66     allocate(seed(n))
67     ! Set seed to a constant so we always get the same matrices
68     seed = 150987
69     call random_seed(put = seed)
70     deallocate(seed)
71
72     ! Allocate host matrices and set up matrices
73     allocate(inputA(rowA, colA))
74     allocate(inputB(rowB, colB))
75     allocate(output(rowA, colB))
76
77     call random_number(inputA)
78     call random_number(inputB)
79
80     ! Allocate and upload device data
81     allocate(d_inputA(rowA, colA))
82     allocate(d_inputB(rowB, colB))
83     allocate(d_output(rowA, colB))
84     d_inputA = inputA
85     d_inputB = inputB
86

```



```

87  blockDim = dim3(block_dim, block_dim, 1)
88  gridDim = dim3((rowA - 1) / block_dim + 1, (colB - 1) / block_dim + 1, 1)
89
90  call mat_mul_kernel <<<gridDim, blockDim>>> (d_inputA, d_inputB, d_output)
91
92  ! Download device data
93  output = d_output
94
95  ! Deallocate variables
96  deallocate(d_inputA)
97  deallocate(d_inputB)
98  deallocate(d_output)
99  deallocate(inputA)
100 deallocate(inputB)
101 deallocate(output)
102 end program main

```

**Listing 2.1:** *Matrix multiplication kernel*

Lines 1 - 43 define a module which contains our matrix multiplication kernel. The kernel takes the matrices `inputA` and `inputB`, multiplies them together, and stores the result in `output`. The kernel has the `attributes(global)` attribute, which means that it can be called from either device or host code. `rowA`, `colA`, `rowB`, and `colB` are the rows and columns of `inputA` and `inputB`. Because they are small and don't change over the course of the calculation, they get stored in a special memory type called constant memory. This is a small cache of read only memory that is 64KB in size and allows for constant values to be read very quickly. There are two shared variables, `s_tileA` and `s_tileB`. Each of these are square matrices with dimensions of `block_dim` and these will be used to store sections or “tiles” of `inputA` and `inputB` respectively. We want each thread to do all the calculations needed to compute a simple element of `output`. Line 13 and 14 use the `threadIdx`, `blockDim`, and `blockIdx` variables so that each thread get a unique combination of `x` and `y` values. `x` corresponds to a row of the `output` matrix and `y` corresponds to a column, so each thread is then uniquely mapped to an element of `output`. We will see in the host code how these index values are determined, but for now it is sufficient to know that there will be at least as many threads as elements of `output`.

Lines 17 - 37 are where the actual calculation takes place. For each iteration of the loop, a different tile of `inputA` and `inputB` are fetched from the global memory and stored into the shared memory. The reason for using this tiling approach can be seen when we consider how the threads must read `inputB`. If we did not use shared memory, the loop would just be `total = total + inputA(x, i) * inputB(i, y)`. In this case, access to `inputA` is coalesced, but access to `inputB` is not. Therefore, the kernel will be stunted due this unoptimized memory usage. Instead, we load tiles in a coalesced manner into the shared memory, synchronize the

threads, and then read through the shared memory in an uncoalesced fashion which carries no penalty. At the end of each loop, the tiles are moved in the **y** direction for **inputA** and in the **x** direction for **inputB**, and the process repeats. Notice how it is possible that a tile might fall on the boundary of an **input** matrix. This is why there are checks to make sure the coordinates of the tile don't exceed the boundaries of a matrix so we don't start reading garbage.

In the host code on lines 45 - 102, we have two sets of variables, those for the host and those for the device. Notice how device variables have the **device** attribute. For clarity, they are given the same name as their host variables but with a **d\_** prefix. Next we assign constant variables from line 5 to arbitrary values. Next we set up the random number generator, allocate the host matrices memory, and randomly initialize the input matrices. Then the device variables are allocated memory and the host data is then uploaded to the GPU. On lines 87 and 88 are where the block and grid dimensions are setup. **blockDim** and **gridDim** are variables of **type(dim3)** which is a tuple that is defined in the **cudafor** module. The block has two dimensions containing **block.dim**<sup>2</sup> threads. The grid also has two dimensions, and as we need one thread per element of **output** we preform a calculation which returns the minimum number of blocks that will cover it. Finally, we launch the kernel on line 90 using the **<<< >>>** operators to provide the grid and block dimensions, download the result from the GPU, and deallocate the memory.

## 2.2 Program flow

In this section, the control flow of a typical calculation is given.

*Step 1.* The input file is read by the **intin** subroutine. If needed, the basis set and open-shell configurations are calculated by **formbs** and **find\_bin\_configurations** respectively. The options set in the input file are rewritten to stdout.

*Step 2.* The calculation of small arrays and other constants is performed by **calc\_parameters** and **bsnorm**. **calc\_parameters** calls **bcoef** and **setvc**.

*Step 3.* The mapping of threads to the unrolled one and two-electron integral matrixes are calculated by **lmpqrsa** on the GPU. All other data from step 1 and 2 is then uploaded to the GPU.

*Step 4.* The one and two electron integrals are calculated on the GPU by **eint1gpu** and **eint2gpu** respectively.

*Step 5.* The initial guess of the density matrix is calculated by **guess** on the GPU. This is done by the diagonalization of the one-electron Hamiltonian matrix.

*Step 6.* **scfiter** then performs the SCF until convergence of the density matrix has been reached, or the maximum number of iterations has been reached. SCF is performed with

cuSOLVER functions, as well as a few custom helper functions. The converged eigenvectors, values, and energies are downloaded from the GPU and then written to stdout.

*Step 7* (optional). If `jobtyp='bsopt'`, then an optional basis set optimization is then carried out. This starts by assigning pointers to variables on the CPU and GPU through *hookup\_cpu* and *hookup\_gpu*. Then, the four WTBS parameters are optimized by *newuoa*. *newuoa* calls *calc\_energy* which essentially reconstructs the basis set from new parameters from *newuoa*, then repeats steps 2 - 6 and feeds the energy back into *newuoa*. This repeats until optimal WTBS parameters have been found.

## 2.3 Alterations for CUDA

While almost all of the code from the original DFRATOM[26] was modified in some way, the most extreme changes were the integral evaluation and the formation of the P and Q matrixes. Therefore, we will discuss these changes in detail in this section.

### 2.3.1 Two-electron Integrals

The original DFRATOM calculates the two electron integrals with a long series of nested **for** loops. Working from the outside in, the indices of the loops are  $L$ ,  $P$ ,  $Q$ ,  $M$ ,  $R$ , and  $S$  where  $L$  and  $M$  are spinor symmetries,  $P$  and  $Q$  are basis functions of symmetry  $L$ , and  $R$  and  $S$  are basis functions of symmetry  $M$ . The pseudocode for these is shown in Algorithm 1. Where  $nsym$  is the total number of symmetries, and  $nbs(i)$  is the number of basis functions for symmetry species  $i$ . Thus, each set of J and K integrals is uniquely defined by its  $L$ ,  $M$ ,  $P$ ,  $Q$ ,  $R$ , and  $S$  values. In order to have an effective CUDA implementation of this algorithm, we need to find a way to map these six numbers to CUDA threads. The simplest approach would be to map the values of  $L$ ,  $P$ , and  $Q$  onto the *threadIdx.x*, *threadIdx.y*, and *threadIdx.z* variables for each thread, launch the needed number of blocks, and then have each thread loop over the remaining indices. The pseudocode for this can be seen in Algorithm 2. A similar method is employed by many other programs, and for larger systems it works perfectly well. But because this program is for only single atoms, problems begin to appear.

The first problem arises due to warp divergence. Because the maximum values of  $M$ ,  $R$ , and  $S$  depend on  $L$ ,  $P$ , and  $Q$ , different threads will have a different number of loops to complete than others. Because a streaming multiprocessor (SM) must finish the block it is currently working on before it can grab another, there is the possibility that most of the threads in a block are idling while waiting for others in the same block to finish. The second problem is that with this method, there will always be several blocks that have threads that

---

**Algorithm 1** The original

---

```
for  $L = 1$  to  $nsym$  do
  for  $P = 1$  to  $nbs(L)$  do
    for  $Q = 1$  to  $P$  do
      for  $M = 1$  to  $L$  do
        if  $L = M$  then
           $maxr = P$ 
        else
           $maxr = nbs(M)$ 
        end if
        for  $R = 1$  to  $maxr$  do
          if  $(L = M)$  and  $(P = R)$  then
             $maxs = Q$ 
          else
             $maxs = R$ 
          end if
          for  $S = 1$  to  $maxs$  do
            compute the J and K integrals of  $L, M, P, Q, R,$  and  $S$ 
          end for
        end for
      end for
    end for
  end for
end for
```

---

---

**Algorithm 2** Easy code

---

$$L = threadIdx.x + (blockIdx.x - 1) * blockDim.x$$
$$P = threadIdx.y + (blockIdx.y - 1) * blockDim.y$$
$$Q = threadIdx.z + (blockIdx.z - 1) * blockDim.z$$

**if**  $(L \leq nsym)$  **and**  $(P \leq nbs(L))$  **and**  $(Q \leq P)$  **then**

**for**  $M = 1$  to  $L$  **do**

**if**  $L = M$  **then**

$maxr = P$

**else**

$maxr = nbs(M)$

**end if**

**for**  $R = 1$  to  $maxr$  **do**

**if**  $(L = M)$  **and**  $(P = R)$  **then**

$maxs = Q$

**else**

$maxs = R$

**end if**

**for**  $S = 1$  to  $maxs$  **do**

        compute the J and K integrals of  $L$ ,  $M$ ,  $P$ ,  $Q$ ,  $R$ , and  $S$

**end for**

**end for**

**end for**

**end if**

---

remain idle throughout the block's runtime, no matter what. This can be seen more clearly in Figure 2.2. The third problem with this is that with the maximum values for  $L$ ,  $P$ , and  $Q$  available for a single atom, there might not even be enough combinations to completely fill the GPU. While all of these issues begin to disappear once the number of integrals to evaluate becomes large enough, we are still very much in the range where they are in play. Therefore a smarter algorithm had to be used.

The second problem can be solved by restricting the dimensionality of the solution instead of the using the three dimensional one of Algorithm 2. We might consider using a two dimensional solution, but then we still will have blocks with idle threads which can be seen in Figure 2.3. By restricting ourselves to only using *threadIdx.x*, we ensure that only the last block to run will have the possibility of having threads remaining idle throughout the block's runtime as shown in Figure 2.4. The third problem can be solved by having each thread calculate one and only one set of J and K integrals. If we start with a valid combination of  $L$ ,  $M$ ,  $P$ ,  $Q$ ,  $R$ , and  $S$  values, we can very easily figure out which thread will calculate that set of integrals by using the following equations:

$$n'(j) = \frac{j^2 + j}{2} \quad (2.1)$$

$$y = n'(nbs(L - 1)) + n'(P - 1) + Q \quad (2.2)$$

$$x = n'(nbs(M - 1)) + n'(R - 1) + S \quad (2.3)$$

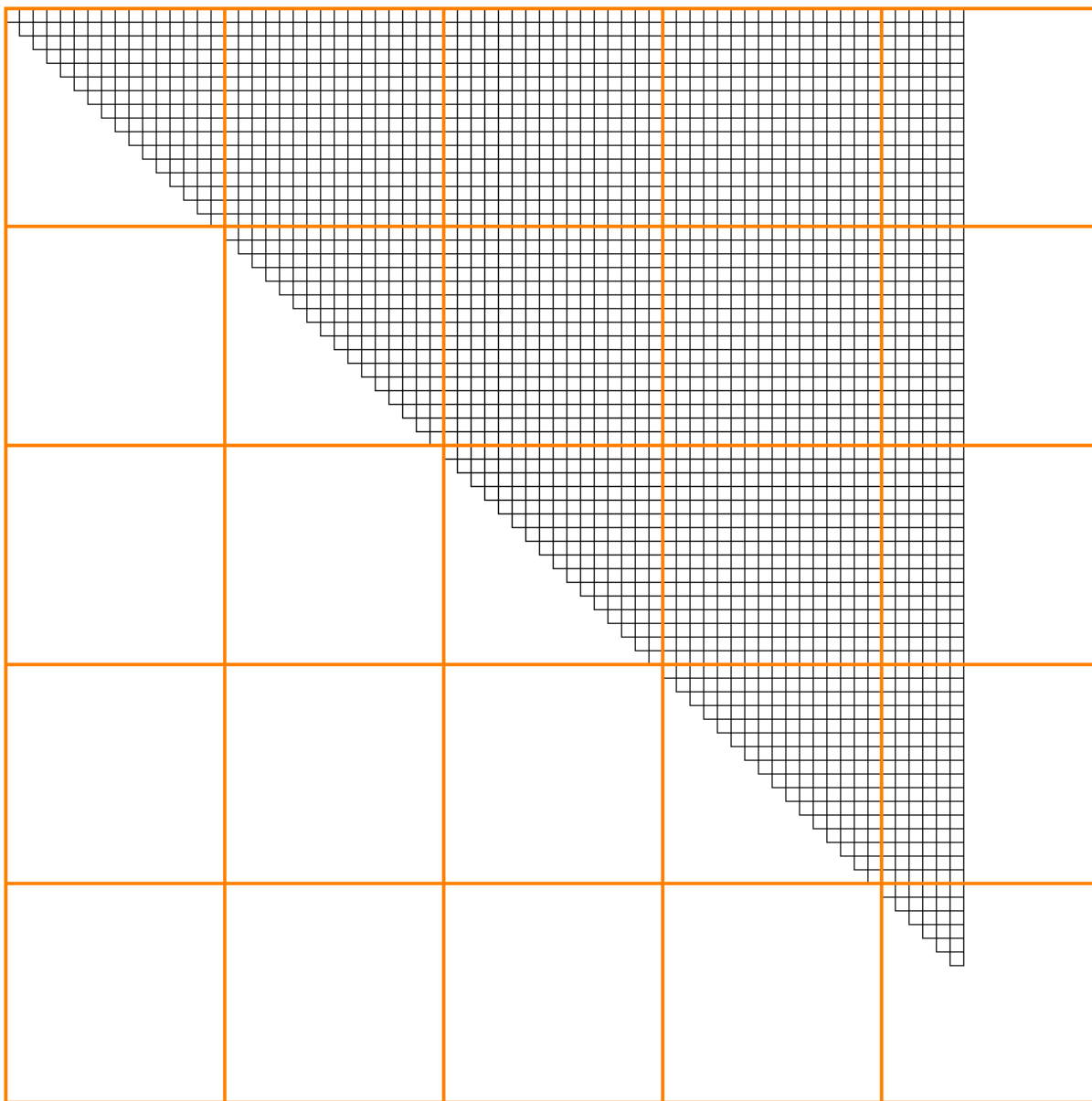
$$i = n'(x) + y \quad (2.4)$$

$$i_{max} = n'(\sum_{j=1}^{nsym} n'(nbs(j))) \quad (2.5)$$

where  $nbs(0) = 0$  and  $i = threadIdx.x + (blockIdx.x - 1) * blockDim.x$ . Starting with a value of  $i$  and working our way back is a much more challenging task. It becomes easier if we reframe it in the following way.

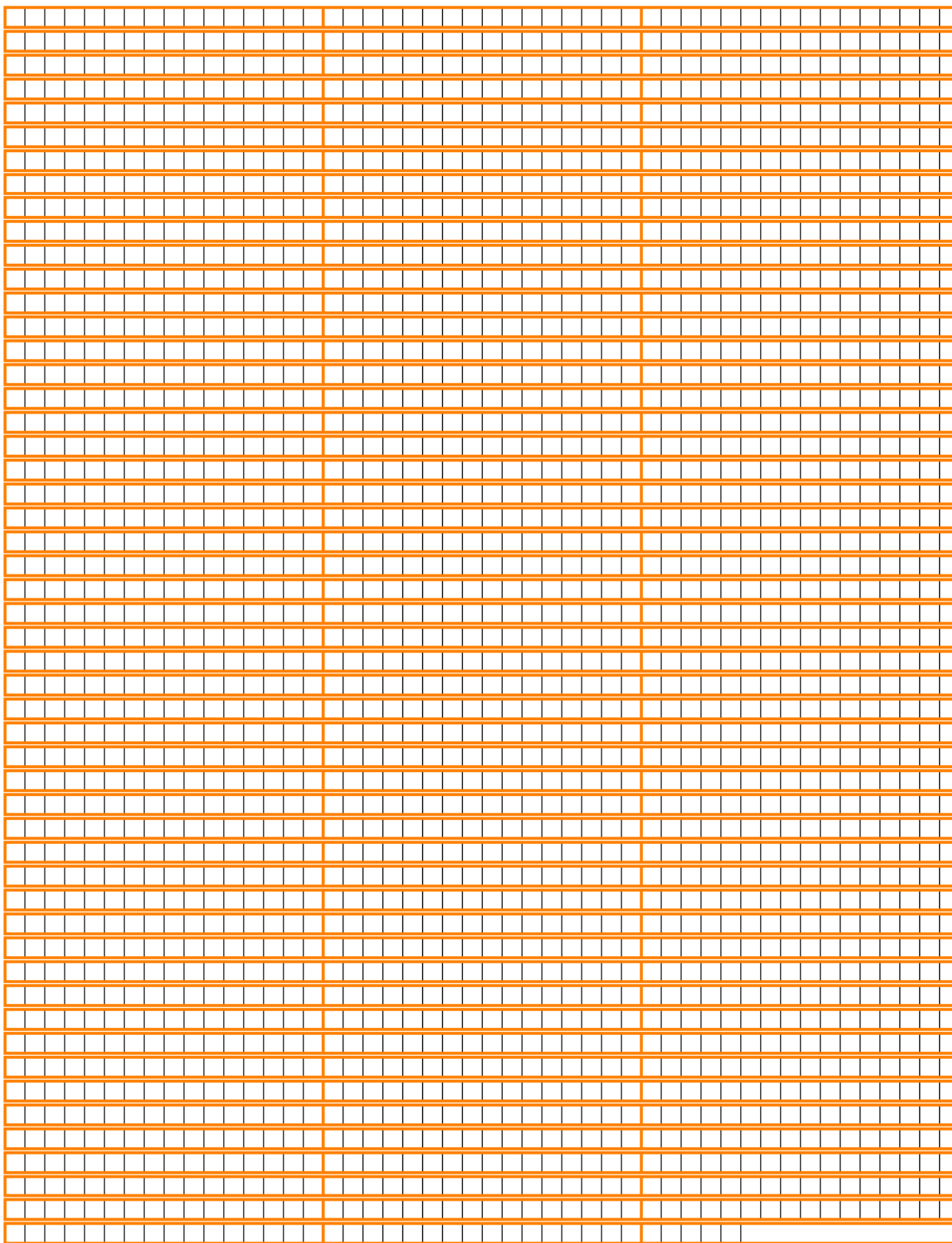


**Figure 2.2:** *Two-electron integral layout: naive implementation.* This figure shows how integrals get mapped to threads using Algorithm 2. The basis set used had  $n_{\text{sym}} = 7$  with  $n_{\text{bs}}(i) = (30, 25, 25, 20, 20, 15, 15)$ . The  $\text{blockDim} = (1, 32, 32)$  and  $\text{gridDim} = (7, 1, 1)$ . Each histogram is one block of threads, where the horizontal axes are the  $\text{threadIdx.y}$  and  $\text{threadIdx.z}$  variables, and the vertical axis is the number of Two-Electron Integrals it must compute. Notice that in all blocks there are threads that calculate no integrals and therefore are idle throughout the lifetime of the block. Also note that the number of integrals each thread must calculate grows both from block to block, and from thread to thread within a block. Both of these factors make this algorithm an undesirable choice.



**Figure 2.3:** *Two-electron integral layout: 2D blocks and grid. Orange lines indicate the border of blocks. Notice that many blocks have idle threads.*





**Figure 2.4:** *Two-electron integral layout: 1D blocks and grid. Orange lines indicate the border of blocks. Notice that only the last block contains idle threads.*

Consider Figures 2.5 and 2.6. Figure 2.5 shows the top half of the symmetric two-electron integral matrix for a problem with  $nsym = 2$  and 3 basis function for symmetry one, and 2 for symmetry two. Figure 2.6 shows the same but mapped onto a vector representation. In each element of both figures, there is a set of seven numbers. The top two are  $L$  and  $M$ , then  $P$  and  $Q$ , then  $R$  and  $S$ , and the last number is the value of  $i$  for the thread calculating that integral. With this, it can be seen from the matrix representation that each element in the same row have the same  $M$ ,  $R$ , and  $S$  values and the elements in the same column have the same  $L$ ,  $P$ , and  $Q$  values. Therefore, finding out which column the element belongs to gives us the value for  $y$ , and finding the row give us the value of  $x$ . This can be done with the binary search algorithm shown in Algorithm 3.

---

**Algorithm 3** Binary search for  $x$  and  $y$

---

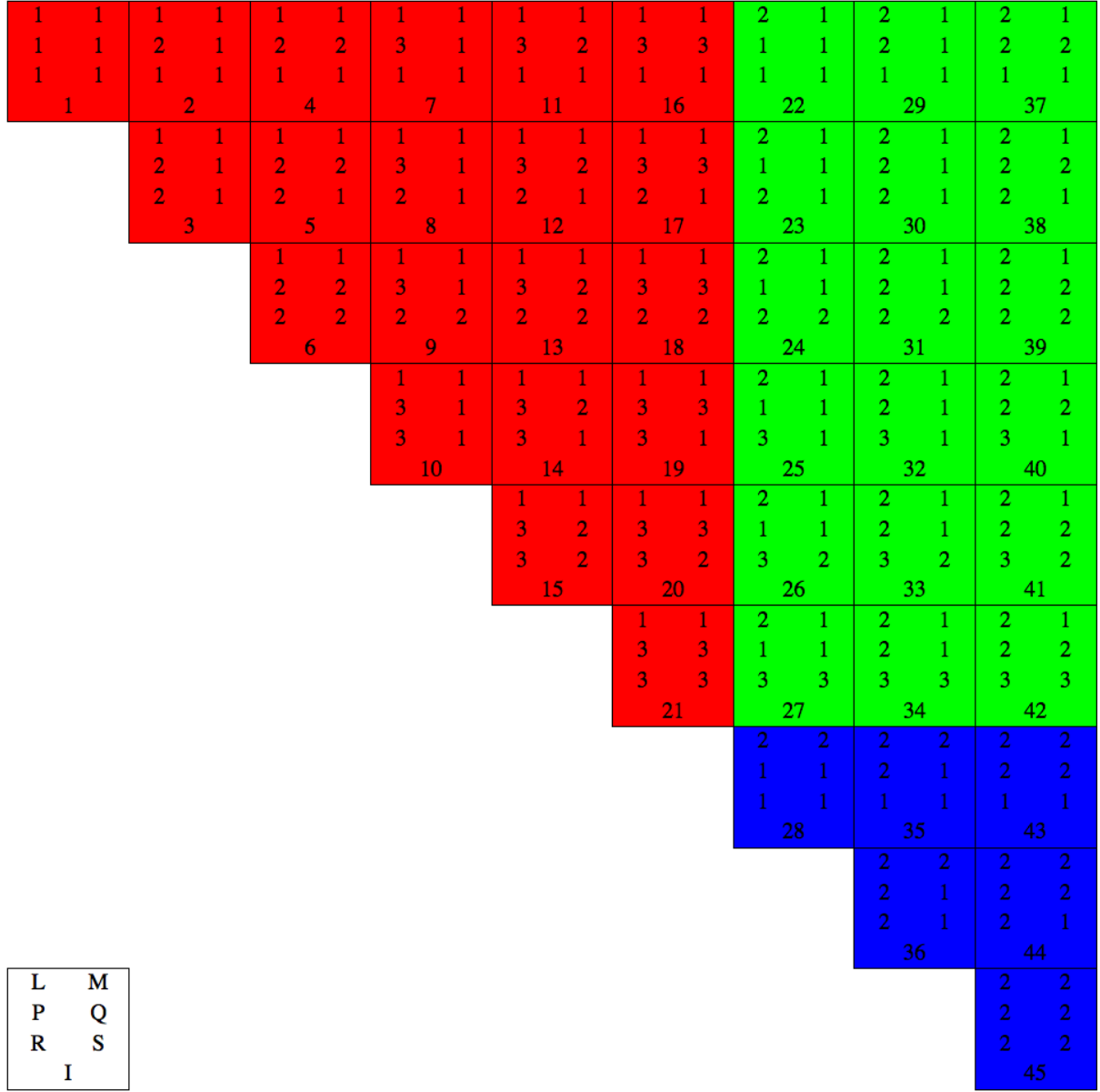
```

if  $theadIdx.x \leq nsym$  then
     $s\_nsym = nsym$ 
     $s\_nbs(theadIdx.x) = nbs(theadIdx.x)$ 
     $s\_nprime(theadIdx.x) = n'(s\_nbs(theadIdx.x))$ 
end if
call syncthreads
 $i = threadIdx.x + (blockIdx.x - 1) * blockDim.x$ 
if  $i \leq i_{max}$  then
     $low = 1$ 
     $high = \text{sum}(s\_nprime(1 : s\_nsym))$ 
    while  $low \leq high$  do
         $mid = \frac{(low+high)}{2}$ 
         $lownum = \frac{(mid-1)(mid-2)}{2} + mid$ 
         $highnum = lownum - 1 + mid$ 
        if  $(i \leq highnum)$  and  $(i \geq lownum)$  then
             $y = mid$ 
            exit
        else if  $i > highnum$  then
             $low = mid + 1$ 
        else if  $i < lownum$  then
             $high = mid - 1$ 
        end if
    end while
     $x = i - lownum + 1$ 
end if

```

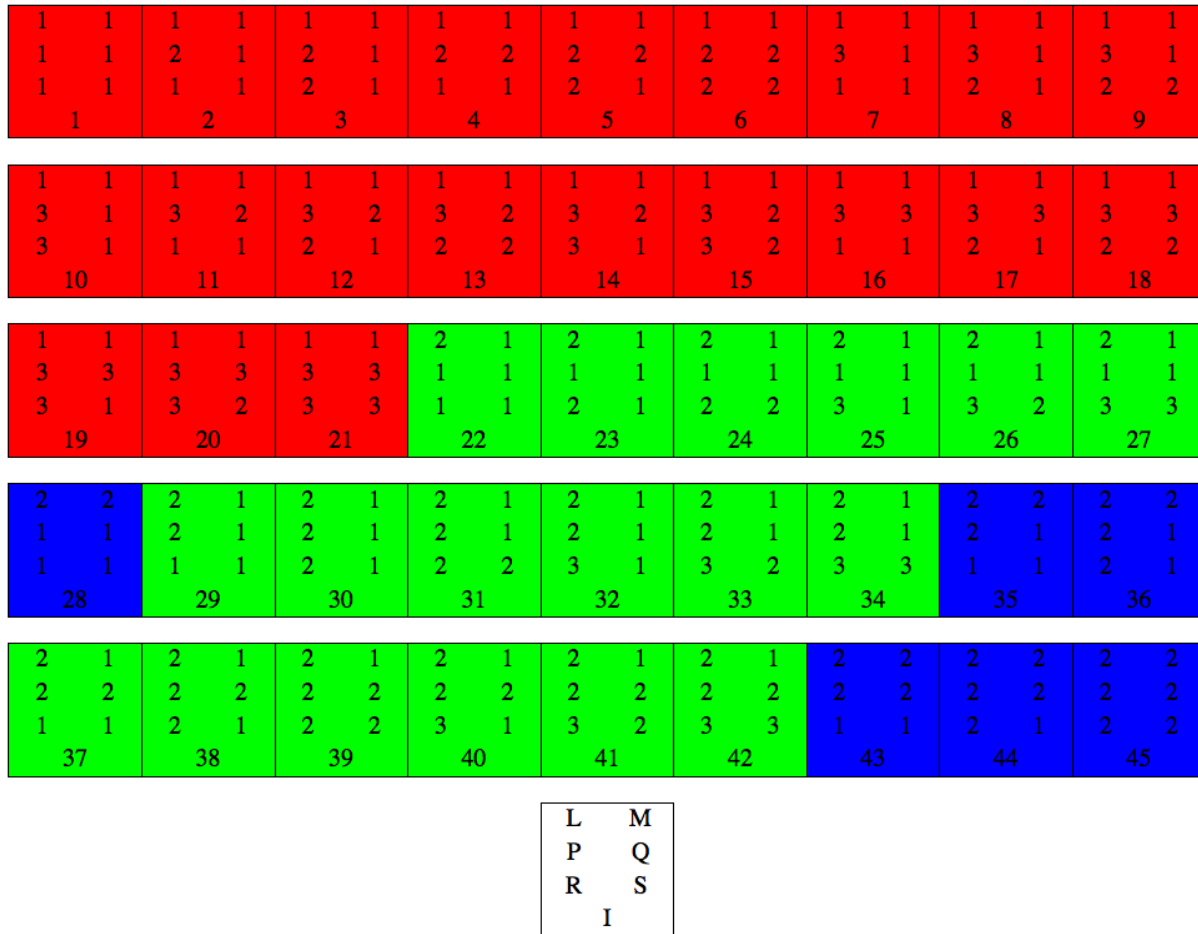
---

In this algorithm, variables with the  $s\_$  prefix refer to those in the shared memory, and  $lownum$  and  $highnum$  refer to the minimum and values the  $i$  could be for the current guess ( $mid$ ) of  $y$ . From here,  $L$  can be found with Algorithm 4, and  $P$  and  $Q$  can be found with Algorithm 5. The same set of algorithms can then be used to get  $M$ ,  $R$ , and  $S$



**Figure 2.5:** Two-electron integral matrix as a matrix. This figure shows the top triangle of the two-electron symmetric matrix in a matrix representation. In this instance, the total number of symmetries is 2, with 3 basis functions for the first symmetry and 2 basis functions for the second. The key for translating the numbers in each box into the L, M, P, Q, R, and S values is given in the bottom left. Notice how all rows have the same M, R, and S values, while all columns have the same L, P, and Q values. For clarity, the boxes have been coloured based on the spinor symmetries that they contain.

by substituting the relevant variables. If there is sufficient global memory available, these values can be stored and referred to later as needed. Otherwise, they could be calculated on the fly as needed. Because binary search scales as  $\mathcal{O}(n \log n)$ , this should ensure that this



**Figure 2.6:** Two-electron integral matrix as a vector. This figure shows the top triangle of the two-electron symmetric matrix in a vector representation. In this instance, the total number of symmetries is 2, with 3 basis functions for the first symmetry and 2 basis functions for the second. The key for translating the numbers in each box into the  $L$ ,  $M$ ,  $P$ ,  $Q$ ,  $R$ , and  $S$  values is given in the bottom center. Notice how the  $L$ ,  $M$ ,  $P$ ,  $Q$ ,  $R$ , and  $S$  values are much harder to predict than from the matrix representation. For clarity, the boxes have been coloured based on the spinor symmetries that they contain. Gaps between rows are used to reinforce that the elements are not connected like a matrix.

remains a fast method of mapping threads to integrals for large problems as well. With some alterations, this method could also apply to molecular symmetries and not only to a single atom. For instance in the  $C_1$  symmetry, all possible combinations of four basis functions must be used (ignoring those that appear on the bottom triangle of the two electron integral matrix of course). We could simply remove the search for  $L$  and  $M$ , have the initial value of *high* in Algorithm 5 be the total number of basis functions, and remove the `sum(s_nprime(1 : L - 1))` term from *lownum*.

From here, the new code for evaluating the integrals remains largely the same as the

---

**Algorithm 4** Binary search for  $L$ 

---

```
 $i = \text{threadIdx}.x + (\text{blockIdx}.x - 1) * \text{blockDim}.x$ 
if  $i \leq i_{\max}$  then
   $low = 1$ 
   $high = s_{\text{nsym}}$ 
  while  $low \leq high$  do
     $mid = \frac{(low+high)}{2}$ 
     $lownum = 1 + \text{sum}(s_{\text{nprime}}(1 : mid - 1))$ 
     $highnum = lownum - 1 + s_{\text{nprime}}(mid)$ 
    if  $(i \leq highnum)$  and  $(i \geq lownum)$  then
       $L = mid$ 
      exit
    else if  $y > highnum$  then
       $low = mid + 1$ 
    else if  $y < lownum$  then
       $high = mid - 1$ 
    end if
  end while
end if
```

---

---

**Algorithm 5** Binary search for  $P$  and  $Q$ 

---

```
 $i = \text{threadIdx}.x + (\text{blockIdx}.x - 1) * \text{blockDim}.x$ 
if  $i \leq i_{\max}$  then
   $low = 1$ 
   $high = s_{\text{nbs}}(L)$ 
  while  $low \leq high$  do
     $mid = \frac{(low+high)}{2}$ 
     $lownum = \frac{(mid-1)(mid-2)}{2} + mid + \text{sum}(s_{\text{nprime}}(1 : L - 1))$ 
     $highnum = lownum + mid - 1$ 
    if  $(i \leq highnum)$  and  $(i \geq lownum)$  then
       $P = mid$ 
      exit
    else if  $y > highnum$  then
       $low = mid + 1$ 
    else if  $y < lownum$  then
       $high = mid - 1$ 
    end if
  end while
   $Q = y - lownum + 1$ 
end if
```

---

original, except for some minor changes to allow for more efficient global or shared memory access. We also use a process referred to as “grid-stride looping” where all these binary search algorithms have their **if**  $i \leq i_{max}$  **then** removed, and then are placed within the following loop: **for**  $i = threadIdx.x + (blockIdx.x - 1) * blockDim.x$  to  $i_{max}$ ,  $i += blockDim.x * gridDim.x$  **do**. If we know the occupancy of the algorithm on the GPU beforehand, we can launch exactly the number of blocks that will fill the GPU. This reduces the overhead of block swapping and lets us further eke out some performance.

### 2.3.2 Density Matrix Formation

The term in parenthesis in Equation 1.35 corresponds to the non-relativistic density matrix. But as we are interested in relativistic calculations, it will need to be altered. Firstly, because a relativistic wavefunction has both large and small components, so too does its density matrix. Therefore, it will have two dimensions that span from 1 to  $2K$  where dimensions 1 to  $K$  are the large components and the rest are for small components. Secondly, the factor of 2 is replaced by the number of electrons that can occupy spinors of symmetry  $\mu$ . Its matrix and vector representations are shown in Figures 2.7 and 2.8 respectively. Its mathematical form is given in Equation 2.6[26]

$$\mathbf{D}_{C_{r,s}} = N_{\mu} \sum_{a=1}^{occ} \mathbf{C}_{r,a} \mathbf{C}_{s,a}^* \quad (2.6)$$

where  $N_i$  is the number of electrons that occupy a closed shell of symmetry  $i$ .  $\mathbf{D}_{\mathbf{O}}$  is similar to this and is shown below.

$$\mathbf{D}_{\mathbf{O}_{r,s}} = N_i \mathbf{C}_{r,(occ+1)} \mathbf{C}_{s,(occ+1)}^* \quad (2.7)$$

where  $N_i$  is equal to the number of electrons in the open shell. Note that in this program, only the most energetic shell of any symmetry can be open.

The total density matrix is  $\mathbf{D}_{\mathbf{T}} = \mathbf{D}_{\mathbf{C}} + \mathbf{D}_{\mathbf{O}}$

#### 2.3.2.1 P Q Super Matrices

With the density matrices in hand we can begin to combine the two-electron integrals from the previous section. These integrals have the following form.[26]

$1^L \quad 1^L$ 1	$2^L \quad 1^L$ 2	$3^L \quad 1^L$ 4	$4^L \quad 1^L$ 7	$5^L \quad 1^L$ 11	$1^S \quad 1^L$ 16	$2^S \quad 1^L$ 22	$3^S \quad 1^L$ 29	$4^S \quad 1^L$ 37	$5^S \quad 1^L$ 46
	$2^L \quad 2^L$ 3	$3^L \quad 2^L$ 5	$4^L \quad 2^L$ 8	$5^L \quad 2^L$ 12	$1^S \quad 2^L$ 17	$2^S \quad 2^L$ 23	$3^S \quad 2^L$ 30	$4^S \quad 2^L$ 38	$5^S \quad 2^L$ 47
		$3^L \quad 3^L$ 6	$4^L \quad 3^L$ 9	$5^L \quad 3^L$ 13	$1^S \quad 3^L$ 18	$2^S \quad 3^L$ 24	$3^S \quad 3^L$ 31	$4^S \quad 3^L$ 39	$5^S \quad 3^L$ 48
			$4^L \quad 4^L$ 10	$5^L \quad 4^L$ 14	$1^S \quad 4^L$ 19	$2^S \quad 4^L$ 25	$3^S \quad 4^L$ 32	$4^S \quad 4^L$ 40	$5^S \quad 4^L$ 49
				$5^L \quad 5^L$ 15	$1^S \quad 5^L$ 20	$2^S \quad 5^L$ 26	$3^S \quad 5^L$ 33	$4^S \quad 5^L$ 41	$5^S \quad 5^L$ 50
					$1^S \quad 1^S$ 21	$2^S \quad 1^S$ 27	$3^S \quad 1^S$ 34	$4^S \quad 1^S$ 42	$5^S \quad 1^S$ 51
						$2^S \quad 2^S$ 28	$3^S \quad 2^S$ 35	$4^S \quad 2^S$ 43	$5^S \quad 2^S$ 52
							$3^S \quad 3^S$ 36	$4^S \quad 3^S$ 44	$5^S \quad 3^S$ 53
								$4^S \quad 4^S$ 45	$5^S \quad 4^S$ 54
									$5^S \quad 5^S$ 55

$R^x$	$S^x$
I	

**Figure 2.7:** *Density matrix as a matrix. This figure shows the top triangle of the symmetric density matrix in a matrix representation. The key for translating the numbers in each box into the  $R$  and  $S$  values is given in the bottom right. All active symmetries would have a similar, separate matrix. For clarity, the elements have been colorized based on the combination of large and small components they contain.*

$1^L$ $1^L$ 1	$2^L$ $1^L$ 2	$2^L$ $2^L$ 3	$3^L$ $1^L$ 4	$3^L$ $2^L$ 5	$3^L$ $3^L$ 6	$4^L$ $1^L$ 7	$4^L$ $2^L$ 8	$4^L$ $3^L$ 9	$4^L$ $4^L$ 10	
$5^L$ $1^L$ 11	$5^L$ $2^L$ 12	$5^L$ $3^L$ 13	$5^L$ $4^L$ 14	$5^L$ $5^L$ 15	$1^S$ $1^L$ 16	$1^S$ $2^L$ 17	$1^S$ $3^L$ 18	$1^S$ $4^L$ 19	$1^S$ $5^L$ 20	
$1^S$ $1^S$ 21	$2^S$ $1^L$ 22	$2^S$ $2^L$ 23	$2^S$ $3^L$ 24	$2^S$ $4^L$ 25	$2^S$ $5^L$ 26	$2^S$ $1^S$ 27	$2^S$ $2^S$ 28	$3^S$ $1^L$ 29	$3^S$ $2^L$ 30	
$3^S$ $3^L$ 31	$3^S$ $4^L$ 32	$3^S$ $5^L$ 33	$3^S$ $1^S$ 34	$3^S$ $2^S$ 35	$3^S$ $3^S$ 36	$4^S$ $1^L$ 37	$4^S$ $2^L$ 38	$4^S$ $3^L$ 39	$4^S$ $4^L$ 40	
$4^S$ $5^L$ 41	$4^S$ $1^S$ 42	$4^S$ $2^S$ 43	$4^S$ $3^S$ 44	$4^S$ $4^S$ 45	$5^S$ $1^L$ 46	$5^S$ $2^L$ 47	$5^S$ $3^L$ 48	$5^S$ $4^L$ 49	$5^S$ $5^L$ 50	
$5^S$ $1^S$ 51	$5^S$ $2^S$ 52	$5^S$ $3^S$ 53	$5^S$ $4^S$ 54	$5^S$ $5^S$ 55						$P^x$ $Q^x$ I

**Figure 2.8:** *Density matrix as a vector. This figure shows the top triangle of the symmetric density matrix in a vector representation. The key for translating the numbers in each box into the  $R$  and  $S$  values is given in the bottom left. All active symmetries would have a similar, similar vector that continue one after the other. For clarity, the elements have been colorized based on the combination of large and small components they contain.*



$$\begin{aligned}
\mathbf{X}_{I(pqrs)}^1 &= (p_L q_L | r_L s_L) - \frac{1}{2} [(p_L r_L | q_L s_L) + (p_L s_L | q_L r_L)] \\
\mathbf{X}_{I(pqrs)}^2 &= (p_L q_L | r_S s_S) \\
\mathbf{X}_{I(pqrs)}^3 &= (p_S q_S | r_L s_L) \\
\mathbf{X}_{I(pqrs)}^4 &= -(p_L r_S | q_S s_L) \\
\mathbf{X}_{I(pqrs)}^5 &= -(p_L s_S | q_S r_L) \\
\mathbf{X}_{I(pqrs)}^6 &= -(p_S r_L | q_L s_S) \\
\mathbf{X}_{I(pqrs)}^7 &= -(p_S s_L | q_L r_S) \\
\mathbf{X}_{I(pqrs)}^8 &= (p_S q_S | r_S s_S) - \frac{1}{2} [(p_S r_S | q_S s_S) + (p_S s_S | q_S r_S)]
\end{aligned} \tag{2.8}$$

where the subscripts  $L$  and  $S$  refer to the large and small components respectively.  $I$  is a function that returns the values of  $x$  and  $y$  from Equations 2.2 and 2.3 where  $L = \lambda$  and  $M = \mu$ . From here, all subscripts of  $\mathbf{X}$  will be implied to go through this  $I$  function.

The  $\mathbf{P}$  and  $\mathbf{Q}$  super matrices have a similar layout to the density matrices. The elements of  $\mathbf{P}$  have the following form[26]

$$\begin{aligned}
\mathbf{P}_{pq}^{LL} &= \sum_{\mu=1}^{nsym} \sum_{r=1} \sum_{s=1} \left( \mathbf{D}_{\mathbf{T}rs}^{LL} \mathbf{X}_{pqrs}^1 + \mathbf{D}_{\mathbf{T}rs}^{SS} \mathbf{X}_{pqrs}^{J(pqrs)} \right) \\
\mathbf{P}_{pq}^{SS} &= \sum_{\mu=1}^{nsym} \sum_{r=1} \sum_{s=1} \left( \mathbf{D}_{\mathbf{T}rs}^{SS} \mathbf{X}_{pqrs}^8 + \mathbf{D}_{\mathbf{T}rs}^{LL} \mathbf{X}_{pqrs}^{J(pqrs)} \right) \\
\mathbf{P}_{pq}^{LS} &= \sum_{\mu=1}^{nsym} \sum_{r=1} \sum_{s=1} \left( \mathbf{D}_{\mathbf{T}rs}^{LS} \mathbf{X}_{pqrs}^5 + \mathbf{D}_{\mathbf{T}rs}^{SL} \mathbf{X}_{pqrs}^{J(pqrs)} \right) \\
\mathbf{P}_{pq}^{SL} &= \sum_{\mu=1}^{nsym} \sum_{r=1} \sum_{s=1} \left( \mathbf{D}_{\mathbf{T}rs}^{SL} \mathbf{X}_{pqrs}^7 + \mathbf{D}_{\mathbf{T}rs}^{LS} \mathbf{X}_{pqrs}^{J(pqrs)} \right)
\end{aligned} \tag{2.9}$$

where  $J(pqrs)$  is an integer function that selects specific definitions of the two-electron integrals from Equation 2.8. How exactly this works is shown in the following sections.

The form of the  $\mathbf{Q}$  super matrix is almost the same as  $\mathbf{P}$ , but there will be an additional multiplication of the vector coupling coefficient between the open shells for each summation.

### 2.3.2.2 FORMPQ Algorithm

Forming the  $\mathbf{P}$  and  $\mathbf{Q}$  matrices proved to be the most difficult part of this program to parallelize. Its difficulty was due to the second term in the summations where each  $\mathbf{X}$  matrix used depends on how far the summation has progressed. Further, each element of

$\mathbf{P}$  will have the  $\mathbf{X}$  matrix change at different times. These two factors result in many **if** statements which GPUs have difficulty handling in parallel. There is also the issue where each element in the  $LL$  and  $SS$  sections need to loop over every element of the same sections of the density matrix, as do the  $LS$  and  $SL$  sections. This means that there will be lots of global memory reads unless handled properly. All together, this makes for a very ugly algorithm to try and parallelize.

Two different attempts were made to try and accelerate this algorithm. Each tried to exploit a different property of GPUs. One attempted to combat the amount of floating point operations to perform by distributing the amount of work to as many threads as possible. This will be referred to as the Multiple Threads Single Element (MTSE) algorithm. The second one tried to deal with the amount of global memory reads by making the most of each read. This will be referred to as Single Thread Single Element (STSE) algorithm. Each will be described in the sections below.

### 2.3.2.3 MTSE Algorithm

As previously stated, the purpose of the MTSE algorithm is to distribute the amount of work needed to be done to as many threads as possible. This was achieved by having multiple threads each computing two of the multiplications needed for a single element of two matrices in Equation 2.9, and then combining the products of these multiplications in parallel. These two tasks were performed by two separate kernels and the pseudocode for each is shown in Algorithms 6 and 7 respectively. The pseudocode shows only the calculation for the

*inttype* refers to one or more if the  $\mathbf{P}$  or  $\mathbf{Q}$  matrices. In Algorithm 6, *inttype* can have a value of 0 or 1, where 0 refers to the  $LL$  and  $SS$   $\mathbf{P}$  or  $\mathbf{Q}$  matrices, and 1 is the  $SL$  and  $LS$  matrices. In Algorithm 7, *inttype* can have a values ranging from 0 to 7 where 0 to 3 are the  $\mathbf{P}$   $LL$ ,  $SS$ ,  $SL$ ,  $LS$  matrices, and the rest are the  $\mathbf{Q}$  matrices in the same order.  $d$  is  $\mathbf{D}$  in its vector representation and  $pos$  are the locations of the  $p$  and  $q$  elements within it. *locmx* is an array with 7 elements, and each element is equal to the number of unique  $LL$ ,  $SS$ ,  $SL$ ,  $LS$  elements of symmetry  $l - 1$ . *fctr* accounts for when an element is single or double counted in the summation.  $x(\chi, :)$  is the vector representation of  $\mathbf{X}^\chi$ , *loc* is the location of needed  $pqr$ s element in the vector, *total* is the length of one of the dimensions of  $\mathbf{X}^\chi$ , and  $k1$  and  $k2$  are the outputs of the  $J$  function from Equation 2.9. Finally,  $pXY$  variables are matrices dimension  $total \times total$  and store the sum of the two multiplications of the  $d$  and  $x$  vectors, *pmx* if the vector representation of the  $\mathbf{P}$  matrices, and *s\_tile* is a vector in the shared memory for summing up the elements of each row of the  $pXY$  matrices.

The grid will need to have  $\frac{total-1}{blockDim.x} \times \frac{total-1}{blockDim.y} \times 2$  dimensions for Algorithm 6 and  $\frac{total-1}{blockDim.x} \times \frac{total-1}{blockDim.y} \times 8$  for Algorithm 7. Algorithm 6 is fairly straightforward, but Algorithm 7 could use a little more explanation. It is a “parallel reduction” algorithm. At the start, each

---

**Algorithm 6** Kernel 1 for MTSE

---

```
i = threadIdx.x + (blockIdx.x - 1) * blockDim.x
j = blockIdx.y
inttype = blockIdx.z - 1
if i ≤ total then
  l = lpq_array(i, 1)
  p = lpq_array(i, 2)
  q = lpq_array(i, 3)
  fctr = 1.0
  if (p ≠ q) and (inttype == 0) then
    fctr = 2.0
  else if (p == q) and (inttype == 1) then
    fctr = 0.5
  end if
  loc = max(i, j) + ((max(i, j) - 1) * (max(i, j) - 2)/2) + min(i, j) - 1
  if (inttype == 0) then
    if i > j then
      k1 = 3
      k2 = 2
    else
      k1 = 2
      k2 = 3
    end if
    pos1 = locmx(l) + ((p - 1)2 + p - 1)/2 + q
    pos2 = locmx(l) + ((p - 1 + nbs(l))2 + p + nbs(l) - 1)/2 + q + nbs(l)
    pLL(i, j) = d(pos1) * fctr * x(1, loc) + d(pos2) * fctr * x(k1, loc)
    pSS(i, j) = d(pos1) * fctr * x(k2, loc) + d(pos2) * fctr * x(k8, loc)
  else
    if i > j then
      k1 = 4
      k2 = 6
    else
      k1 = 6
      k2 = 4
    end if
    pos1 = locmx(l) + ((p - 1 + nbs(l))2 + p - 1 + nbs(l))/2 + q
    pos2 = locmx(l) + ((q - 1 + nbs(l))2 + q - 1 + nbs(l))/2 + p
    pSL(i, j) = d(pos1) * fctr * x(7, loc) + d(pos2) * fctr * x(k1, loc)
    pLS(i, j) = d(pos1) * fctr * x(k2, loc) + d(pos2) * fctr * x(5, loc)
  end if
end if
```

---

---

**Algorithm 7** Kernel 2 for MTSE

---

```
i = threadIdx.x + (blockIdx.x - 1) * blockDim.x * 2
inttype = blockIdx.z - 1
if inttype == 0 then
  if i ≤ total then
    s_tile(threadIdx.x) = pLL(i, blockIdx.y)
  else
    s_tile(threadIdx.x) = 0.0
  end if
  if i + blockDim.x ≤ total then
    s_tile(threadIdx.x + blockDim.x) = pLL(i + blockDim.x, blockIdx.y)
  else
    s_tile(threadIdx.x + blockDim.x) = 0.0
  end if
else
  Perform a similar instruction on the other pXY or qXY matrices depending on inttype
end if
for j =  $\log_2(\text{blockDim.x})$ , j ≥ 0, j = j - 1 do
  call syncthreads
  if threadIdx.x ≤  $2^j$  then
    s_tile(threadIdx.x) = s_tile(threadIdx.x) + s_tile(threadIdx.x +  $2^j$ )
  end if
end for
if threadIdx.x == 1 then
  l = lpq_array(blockIdx.y, 1)
  p = lpq_array(blockIdx.y, 2)
  q = lpq_array(blockIdx.y, 3)
  if (inttype == 0) then
    pos = locmx(l) + ((p - 1)2 + p - 1)/2 + q
    istat = atomicadd(pmx(pos), s_tile(threadIdx.x))
  else
    Perform a similar instruction on the other pXY or qXY matrices depending on inttype
  end if
end if
```

---

thread loads an element of the relevant  $pXY$  array into the shared variable  $s\_tile$  and also loads the element that is  $blockDim.x$  elements ahead of the original element. Then, in a loop starting from  $j = \log_2(blockDim.x)$  and descending to  $j = 0$  threads with a  $threadIdx.x$  value of less than  $2^j$  add the value of  $s\_tile(threadIdx.x + 2^j)$  to  $s\_tile(threadIdx.x)$ . The key feature is that the amount of elements that need to be added up is reduced by half every loop. Because it is possible that the total number of elements of  $pXY$  needed to sum up for the relevant  $pmx$  element could be larger than one block can handle, atomic operations are used to force the final addition to occur serially.

### 2.3.2.4 STSE Algorithm

The STSE algorithm tries a different approach to this problem than the MTSE algorithm. Instead of trying to spread out the amount of work to do as much as possible, it tries to reduce the amount of global memory reads by making each thread do as much work as it can. This results in fewer blocks, but each block needs to run longer. Giving only the calculations for  $pmx$  and when  $inttype = 0$ , Algorithm 8 shows how this is done.

All variables are the same as in Algorithms 6 and 7, but  $inttype$  can only be 0 or 1 (the meanings are the same as in Algorithm 6). As for the new variables,  $s\_d$  is a chunk of  $d$  that has been loaded into the shared memory.  $minx$  and  $maxx$  are used to calculate the sections of  $x$  that current chunk of  $d$  in shared memory applies to and  $loop$  is used to adjust the value of  $loc$  if the needed two-electron integral is on the bottom triangle of a  $\mathbf{X}$  matrix.  $pm1$  and  $pm2$  are accumulator variables that track the sum of an element of  $pmx$  as  $d$  is looped through.

Algorithm 8 has two nested loops. The outer one loops over elements of  $d$ , loading chunks of it into the shared memory as it goes. The inner one loops over elements of  $x$  and performs the multiplications of  $d$  and  $x$  elements, keeping a running tally of the sum of these multiplications. Something that is not shown in the pseudocode is checking if  $l$  and  $m$  have open shells. If both of them do, then  $qm1$  and  $qm2$  will be calculated as well. If only one or neither of them do, then this calculation is skipped. Because most of the shells will be closed, this allows for many calculations to be skipped. Once the inner loop has finished, the next section of  $d$  is loaded into the shared memory and then the multiplication of  $x$  and  $s\_d$  continues. This process goes on until all elements of a row of  $\mathbf{X}$  have been used.

## 2.4 Discussion

In this section, the results of GPU resource usage and profiling the various algorithms discussed previously will be presented. For profiling, the radon atom was chosen. Radon was selected because it contains many spinor orbitals and should be a good test of the *formpq*

---

**Algorithm 8** STSE algorithm

---

```
i = threadIdx.x + (blockIdx.x - 1) * blockDim.x
inttype = blockIdx.z - 1
if i ≤ total then
  l = lpq_array(i, 1)
  p = lpq_array(i, 2)
  q = lpq_array(i, 3)
  fctr = 1.0
  loop = 0
  for j = 1, j ≤ (total - 1)/blockDim.x + 1, j = j + 1 do
    mink = 1 + (j - 1) * blockDim.x
    maxk = min(j * blockDim.x, total)
    if threadIdx.x + (j - 1) * blockDim.x ≤ total then
      m = lpq_array(threadIdx.x + (j - 1) * blockDim.x, 1)
      r = lpq_array(threadIdx.x + (j - 1) * blockDim.x, 2)
      s = lpq_array(threadIdx.x + (j - 1) * blockDim.x, 3)
      if (r ≠ s) and (inttype == 0) then
        fctr = 2.0
      end if
      s_d(threadIdx.x, 1) = fctr * d(locmx(m) + ((r - 1)2 + r - 1)/2 + s - 1)
      s_d(threadIdx.x, 2) = fctr * d(locmx(m) + ((r - 1 + nbs(m))2 + r + nbs(m) - 1)/2 +
        s + nbs(m))
    end if
    fctr = 1.0
    call syncthreads
    for k = mink, k ≤ maxk, k = k + 1 do
      if (i ≥ k) and (inttype == 0) then
        loc = ((i - 1)2 + i - 1)/2 + k
        k1 = 3
        k2 = 2
      else if (i < k) and (inttype == 0) then
        loop = loop + k - 1
        loc = (i2 + i)/2 + loop
        k1 = 2
        k2 = 3
      end if
      pm1 = pm1 + s_d(1 + k - mink, 1) * x(1, loc) + s_d(1 + k - mink, 2) * x(k1, loc)
      pm2 = pm2 + s_d(1 + k - mink, 1) * x(k2, loc) + s_d(1 + k - mink, 2) * x(k8, loc)
    end for
    call syncthreads
  end for
  pmx(locmx(l) + ((p - 1)2 + p - 1)/2 + q) = pm1
  pmx(locmx(l) + ((p - 1 + nbs(l))2 + p + nbs(l) - 1)/2 + q + nbs(l)) = pm2
end if
```

---

and *eint2* kernels on large atomic systems. Here, we only care about how long a calculation will take, not how accurate it is. Therefore, basis sets use the arbitrarily chosen WTBS parameters of  $\alpha = 6.87 \times 10^{-2}$ ,  $\beta = 1.83$ ,  $\delta = 3.97$ , and  $\gamma = 0.78$ . Additionally, due to the spherical symmetry of an atom exploited by cudaDFRATOM, the growth of a calculation differs depending on how the basis functions are distributed among the spinor symmetries. Therefore, results are reported in terms of the number of two-electron integrals instead of number of basis functions as is typically done.

### 2.4.1 Resource Usage

In the case of the work in this Thesis, the resource bottleneck for my kernels was almost always the number of registers used per thread. 63 32-bit registers provide only 252 bytes to work with, which when using double precision 8-byte numbers is hardly anything at all. A summary of the block sizes and resource usage of all kernels is given in Table 2.1.

**Table 2.1:** *Kernel resource usage*

Kernel	Block Size	Registers	Shared Memory (bytes)
<i>eint1</i>	(512, 1, 1)	63	45056
<i>eint2</i>	(512, 1, 1)	63	45088
<i>vec2matrix</i>	(512, 1, 1)	17	0
<i>formd</i>	(512, 1, 1)	33	28672
<i>binary_search</i>	(512, 1, 1)	30	60
<i>convd</i>	(512, 1, 1)	25	0
<i>formpq_alg1</i>	(512, 1, 1)	46	32768
<i>formpq_alg2</i>	(512, 1, 1)	56	39388
<i>forme</i>	(512, 1, 1)	39	4152
<i>formf</i>	(512, 1, 1)	18	0
<i>xtrpf</i>	(512, 1, 1)	12	0
<i>formg</i>	(32, 16, 1)	55	12288
<i>swapcol</i>	(32, 16, 1)	10	0

Of these kernels, *eint2* and the *formpq\_alg* kernels take up the bulk of the time needed in a typical calculation. This is in part due to the exponential growth in the number of two-electron integrals, but it is also caused by the amount of register spilling that each kernel will have. As stated in section 2.1.2 when the amount of registers a thread needs exceeds the amount that is available, it will begin to store the data in the global memory. As global memory is the slowest available memory type, this is very undesirable. To counteract this, many variables that would otherwise be better stored in register memory were moved into the shared memory. But this is also undesirable because it means that there will be less shared

memory available for storing data that needs to be read frequently from global memory. This in turn leads to more frequent reads from the global memory. Solving this balancing act required many iterations of testing, moving variables around, compiling, and retesting. Eventually, I arrived at the configurations shown above. Note how each kernel uses a block containing 512 threads. This number was chosen because it is the largest possible block size that can run when the required number of register per thread exceeds 32. 512 is also a good block size because in instances where a kernel is not limited by register or shared memory usage, it allows for the maximum number of threads per SM (1536) to be run.

### 2.4.2 Two-Electron Integral performance

Table 2.2 and Figure 2.9 show the results of profiling the *eint2* kernel against its CPU counterpart.

**Table 2.2:** *eint2* kernel performance. Times are in ms.

Two-Electron Integrals <sup>a</sup>	CPU Time	<i>eint2</i>	<i>binary_search</i>	Total GPU time	Speedup <sup>b</sup>	Speedup <sup>c</sup>
74305	59	4.68	0.18	4.86	12.81	12.32
353220	320	18.05	0.71	18.77	17.72	17.04
1081185	980	52.51	2.21	54.72	18.66	17.90
2588950	2360	118.20	5.36	123.56	19.96	19.09
5299140	4910	237.96	11.07	249.03	20.63	19.71
9726255	8880	437.88	20.81	458.69	20.27	19.35
16476670	15140	730.24	35.96	766.20	20.73	19.75
26248635	24180	1159.98	57.99	1217.97	20.84	19.85

<sup>a</sup> For only the upper triangle of one of the 8 **X** matrices.

<sup>b</sup> Speedup comparing just the time of *eint2* to the CPU time.

<sup>c</sup> Speedup comparing the time of *eint2* and *binary\_search* to the CPU time.

As can be seen, the GPU calculations outperform the CPU calculations with an average speedup factor of over 19. For calculations with fewer number of integrals to compute the speedup is not as dramatic. This is most likely due to the overhead involved in just starting a GPU kernel. Table 2.2 shows two different speedups: one for just calculation of the integrals, and one that shows the speedup if the time needed for *binary\_search* is included. The time needed to evaluate indices does not have much of an impact on the total performance on the total calculation time which proves this was a good way to generate the *lmpqrs* indices. Additionally, because cudaDFRATOM is a basis set optimizer, *binary\_search* will need to





**Figure 2.9:** *Profiling of two-electron Integral calculations.* This figure shows the results of profiling the calculation of growing amounts of two-electron integrals. The blue line shows the time needed for GPU calculations and the yellow line shows time needed for CPU calculations. The time for the GPU calculations is the combined time of the *eint2* and *binary\_search* kernels.

be called only once, whereas *eint2* may need to be called many hundreds of times. Therefore, for an optimization calculation (which the original program is incapable of doing) the total speedup will limit to the just the integral calculations.

### 2.4.3 P and Q Matrix Formation Performance

As explained in section 2.3.2.1, there were two different attempts to speedup the formation of **P** and **Q**. The results of these attempts are shown in Tables 2.3 and 2.4 and also in Figure 2.10.

As can be seen, the attempt at *formpq\_alg1* was a complete failure and it isn't even faster than the CPU calculations. *formpq\_alg2* was much better, but only just. It achieves little more than a two-fold speedup at best.

There are several reasons this could be the case. The first due to the poor global memory access that reading *x* has. *x* is stored as a vector. This helps greatly with memory cost and integral calculation, but makes it difficult to retrieve it as if it where a matrix. There is the

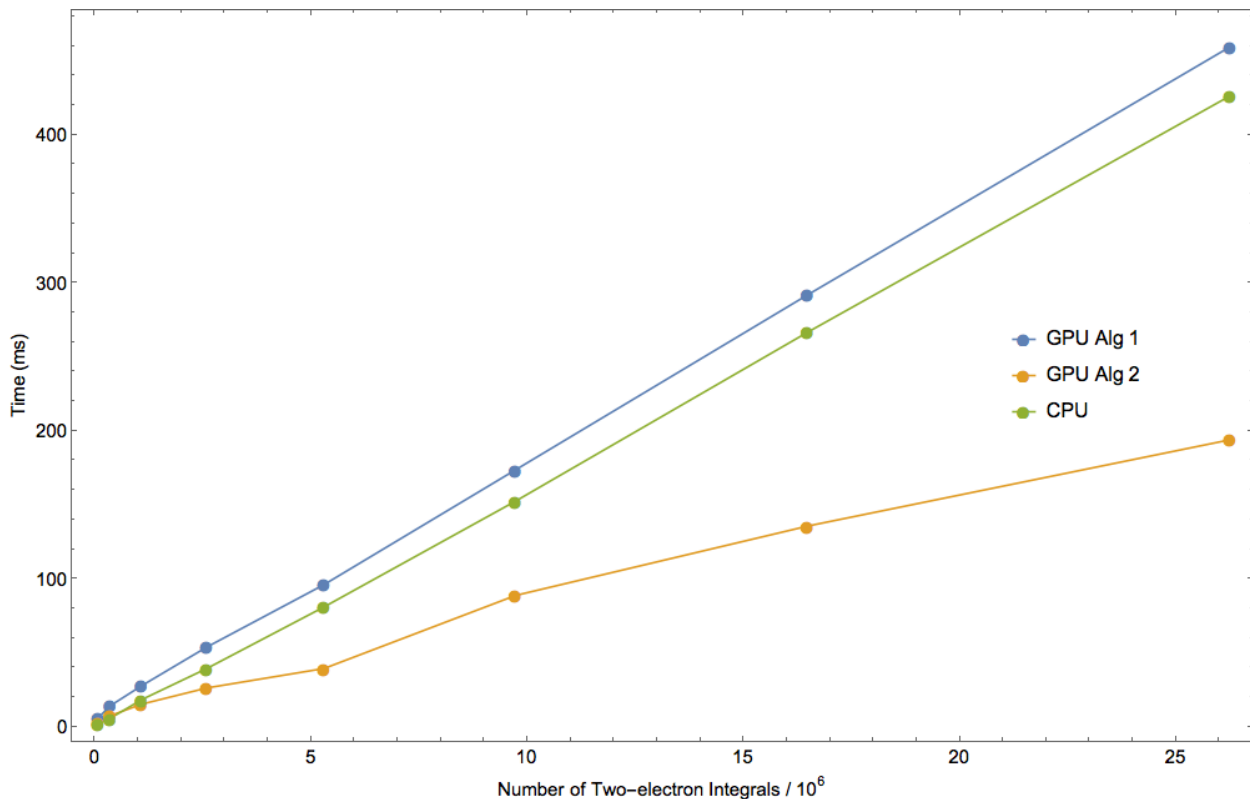
**Table 2.3:** *formpq\_alg1* kernel profiling results. Times are in ms.

Two-Electron Integrals	Time per Iteration CPU	Time per Iteration GPU	Speedup
74305	1.90	6.06	0.31
353220	5.78	14.25	0.40
1081185	18.42	27.81	0.66
2588950	39.47	53.95	0.73
5299140	81.05	96.13	0.84
9726255	152.50	173.64	0.87
16476670	266.53	291.79	0.91
26248635	426.25	459.38	0.92

**Table 2.4:** *formpq\_alg2* kernel profiling results. Times are in ms.

Two-Electron Integrals	Time per Iteration CPU	Time per Iteration GPU	Speedup
74305	1.90	2.94	0.64
353220	5.78	7.56	0.76
1081185	18.42	15.34	1.20
2588950	39.47	26.32	1.49
5299140	81.05	39.45	2.05
9726255	152.50	88.78	1.71
16476670	266.53	135.64	1.96
26248635	426.25	193.99	2.19

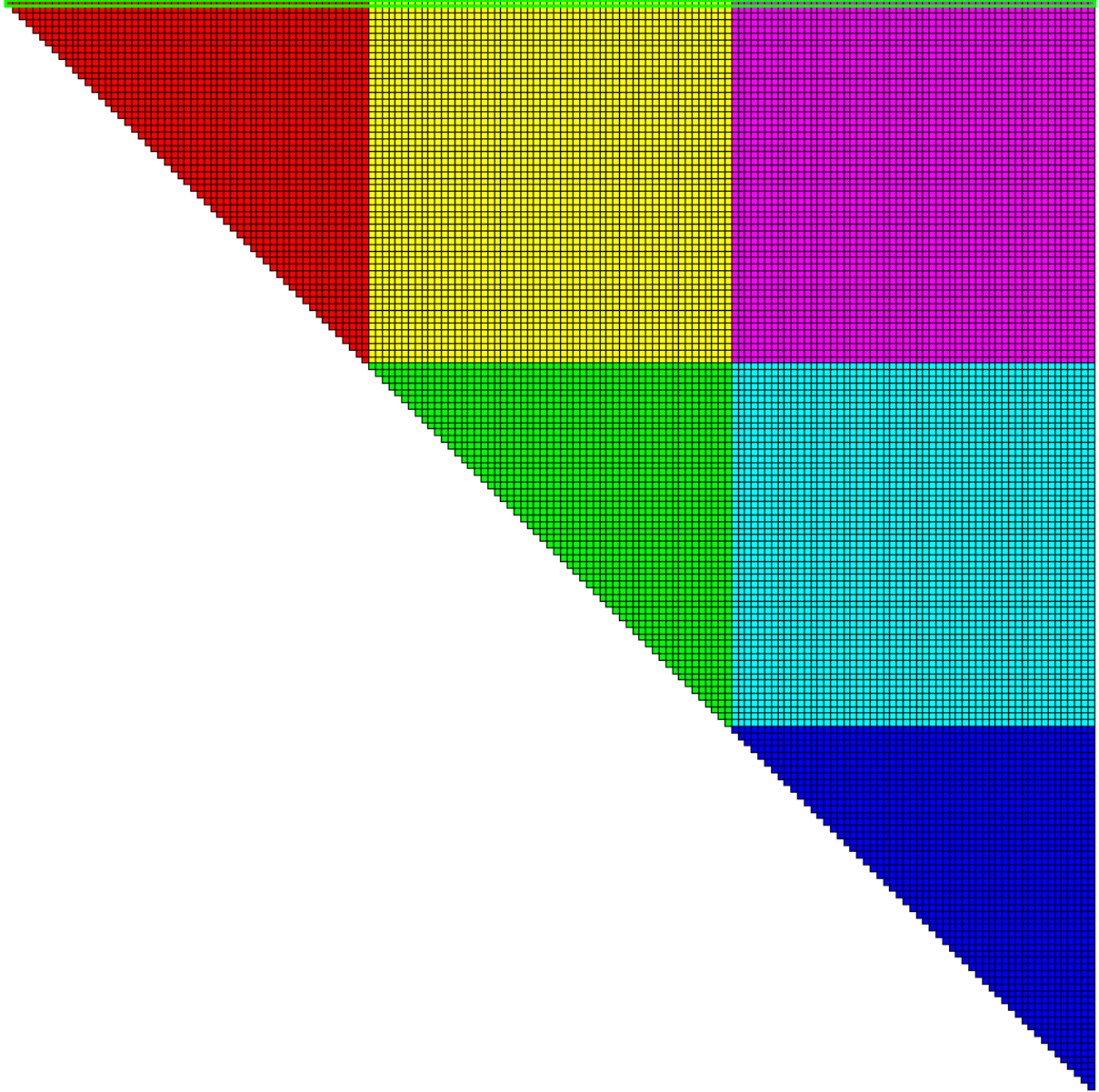
cost of calculating where the necessary integral is, and then the cost of the very uncoalesced reads. Calculating the integrals as a vector, and then copying them into a matrix afterwards could help with this, but the large amount of integrals needed for large calculations makes this impractical on current hardware. How much of an effect inefficient access of global memory has on a calculation does have a limit however which is most likely why the speedup sees an improvement as the calculations get larger. A second reason that the speedup is limited is caused by the occupancy that the kernels can achieve on the GPU. A block size of 512 means that there is a maximum number of 3 blocks running per SM at a time. But these kernels are both limited by the register usage to just running one block per SM at a time. This is also most likely why *formpq\_alg1* performs so poorly compared *formpq\_alg2*. *formpq\_alg1* was designed to spread out the amount of work to as many threads as possible. This results in a lot of blocks to be run. But because the amount of blocks that can run at a time is so limited there is no chance to see an improvement. But in the case of *formpq\_alg2*, there are a fewer number of blocks needing be be run, although each one needs to run for longer



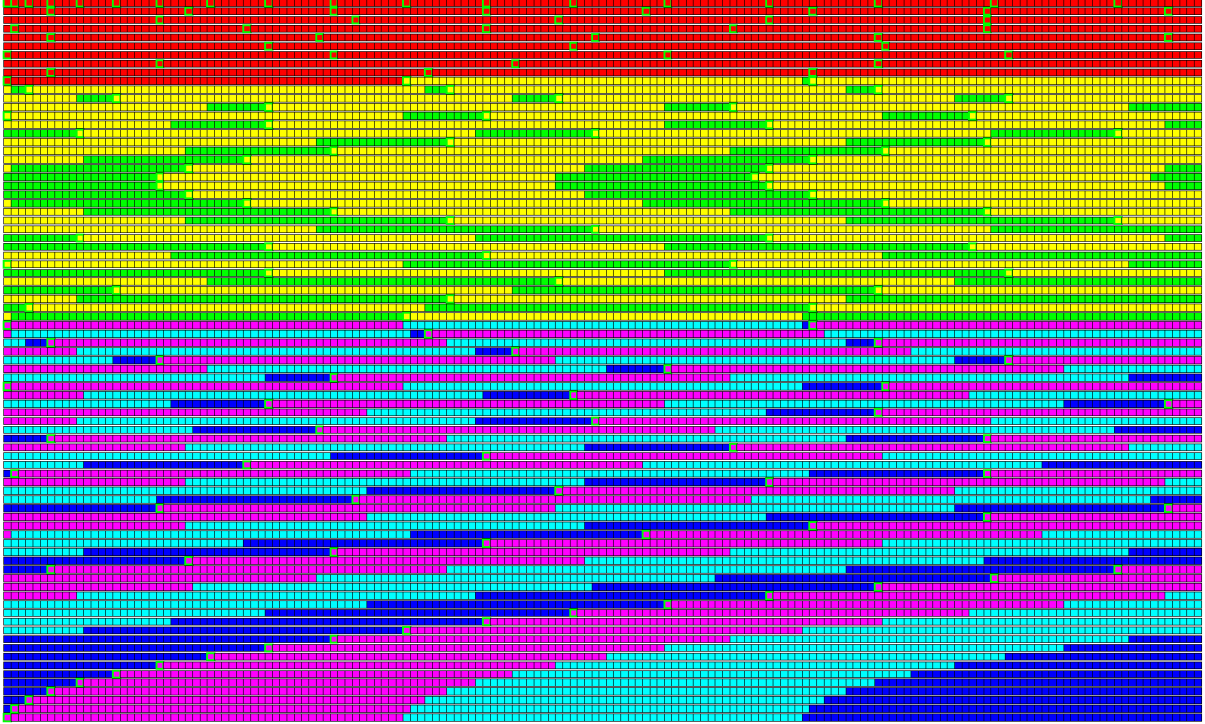
**Figure 2.10:** Profiling of the **formpq** kernels. This figure shows the results of profiling the formation of the **P** and **Q** matrices. The blue line shows the time needed for GPU calculations using the **formpq\_alg1** algorithm, the yellow line shows time needed for GPU calculations using **formpq\_alg2**, and the green line shows the time needed for the CPU. Because the GPU calculations converge in fewer SCF iterations than the CPU calculations, only the average time for one iteration is shown.

than those of **formpq\_alg1**. Overall **formpq\_alg1** comes out on top.

How long will this be true though? At the moment, the limitations of these algorithms are caused by the hardware they run on, not the efficiency of the algorithms themselves. If we imagine removing these limitations, we could consider which one will be better. I propose that **formpq\_alg1** will eventually surpass **formpq\_alg2** for very large systems for the following reasons: the time it takes for a single block of **formpq\_alg1** to run is mostly governed by how well it can read the  $x$  vector. The effect of spreading out the elements has on memory bandwidth has a limit, once the elements are spread out enough, spreading them out further has no impact. Therefore there is a maximum amount of time it takes for a single block to run. The time for a single block of **formpq\_alg1** on the other hand is controlled by both the efficiency of reading global memory *and* the amount of two-electron integrals it must read though. So there is no limit to how long a block can take. Therefore, I predict that **formpq\_alg1** should be faster as the number of integrals becomes very large.



**Figure 2.11:** Integrals needed for the first element of  $\mathbf{P}$ : matrix representation. This figure shows the top triangle of the symmetric two-electron integral matrix in a matrix representation. The required integrals have been highlighted with a green boarder. The basis set used had  $nsym = 3$  with  $nbs(i) = (10, 10, 10)$ . For clarity, the boxes have been coloured based on the spinor symmetries that they contain.



**Figure 2.12:** Integrals needed for the first element of  $\mathbf{P}$ : vector representation. This figure shows the top triangle of the symmetric two-electron integral matrix in a vector representation. The required integrals have been highlighted with a green boarder. The spreading out of the elements means that reads to them can not be done effectively. The basis set used had  $n_{\text{sym}} = 3$  with  $n_{\text{bs}}(i) = (10, 10, 10)$ . For clarity, the boxes have been coloured in the same way as Figure 2.12.

## 2.4.4 Total Speedup

There was some difficulty in determining the total speedup of a calculation. The reason why can be seen in Tables 2.5 and 2.6.

**Table 2.5:** *Total speedup of alg1. The results for a total calculation are shown. All times are in s.*

Two-Electron Integrals	SCF Iterations CPU	Total CPU Time	SCF Iteractions GPU	Total GPU Time	Speedup
74305	21	0.12	21	0.84	0.14
353220	19	0.49	19	1.19	0.41
1081185	19	1.44	19	1.75	0.82
2588950	19	3.38	19	2.58	1.30
5299140	19	6.78	19	3.80	1.78
9726255	24	13.29	19	5.77	2.30
16476670	26	23.23	19	8.60	2.69
26248635	24	35.63	19	12.5	2.82

**Table 2.6:** *Total speedup of alg2. The results for a total calculation are shown. All times are in s.*

Two-Electron Integrals	SCF Iterations CPU	Total CPU Time	SCF Iteractions GPU	Total GPU Time	Speedup
74305	21	0.12	21	0.78	0.15
353220	19	0.49	19	1.07	0.46
1081185	19	1.44	19	1.51	0.95
2588950	19	3.38	19	2.08	1.62
5299140	19	6.78	19	2.72	2.49
9726255	24	13.29	19	4.14	3.20
16476670	26	23.23	19	5.57	4.16
26248635	24	35.63	19	7.47	4.76

As can be seen, the number of SCF iterations differs when comparing the GPU calculations to the CPU even though the convergence criteria was exactly the same. This is somewhat puzzling but I offer the following explanation. First, when writing the SCF routines, it became apparent that rewriting the linear algebra into CUDA would be rather pointless when I could just use the cuSOLVER and cuBLAS libraries and achieve the same result. As such, the SCF is completed using entirely different algorithms. Second, it is important to realize that a real number can only be *approximated* using double precision variables.



**Figure 2.13:** *Total time need to complete a calculation.*

This introduces numerical instability. For instance, assume that we have a computer that could only hold the first four digits of a number (leading zeros don't count). Then we take the number 1.000 and we want to add 0.0001000 to it 10000 times. The result of the first addition should be 1.0001, but because our computer can only hold four digits, the number is truncated to 1.000. The final result of this computation would not be 2.000 as expected but 1.000. During an SCF calculation, there are millions of numbers being multiplied together and added up, many of which may be close to, but not quite 0. This allows for lots of error to creep into a calculation just because of the limitations of the hardware. The fact that two different algorithms are being used on the GPU and CPU, and that they might not be handling numerical stability in the same way is what is most likely causing this to occur. It should also be pointed out that the final energy was not exactly the same from GPU to CPU either. Typically, the last 4 of 16 significant digits would be different. Because the rest are the same however, we still have chemical accuracy and I argue that cudaDFRATOMs results are therefore valid.

## 2.5 Input Description

The program can be executed on Unix-like systems in the following way

```
$ path_to_executable input_file > output_file
```

The input file must end in “.inp” or an error will be given. Redirection of stdout to an output file is optional, but is recommended to save the results of a calculation. The input file is read using the namelist functionality of Fortran. A description of what must appear on each line of the input file is given below. Sample input files are also given in [Appendix whatever](#).

1. A title of no more than 200 characters.

2. \$contrl

jobtype		The type of calculation to be performed.
=	'energy'	Will do a single point energy calculation.
=	'bsopt'	Will optimize the basis set. Can only be used if bastype equals 'WTBS'.
c		The speed of light. If not given, the default is set to 137.03599976 au.

3. \$nuc

znuc		The charge of the nucleus.
nucmdl		The nuclear model to use.
=	1	Point nucleus (default).
=	2	Finite sphere (not yet supported).
=	3	Gaussian.
rnuc		The radius of the finite sphere nucleus.
alpha		The exponent for the gaussian nucleus.
		Defaults are given in litdata.f90.

4. \$bas

nsym		The number of symmetries to be used.
bastype		The type of basis set given.
=	'WTBS'	Use a WTBS.
=	'rdin'	Read in the basis set from the input file.
ngroup		The number of different groups to use in the WTBS scheme (default 1).

The next line will depend on the value of bastype. If bastype equals 'rdin' then the following lines must be the number of functions for the S+ symmetry, followed by the



exponents to use, each on a new line. The pattern repeats for each new symmetry. See the sample input files for further clarification. Otherwise, if `bastype` equals ‘WTBS’ the \$WTBS group is read next.

5. \$WTBS has to be given if `bastype`=‘WTBS’.

WTBSpara	The $\alpha$ , $\beta$ , $\delta$ , and $\gamma$ WTBS parameters. If there is more than one group, the order would be $\alpha_1$ , $\beta_1$ , $\delta_1$ , $\gamma_1$ , $\alpha_2$ , $\beta_2$ , $\delta_2$ , $\gamma_2$ and so on.
nbs	The number of functions used in each symmetry.
start	Where in the $\zeta$ pool each symmetry starts taking exponents from (default=1,1,1,1,1,1,1).
groups	What group each symmetry belongs to (default=1,1,1,1,1,1,1).

The next line depends on what the `jobtype` was set to. If `jobtype` equals ‘energy’ \$newuoa is skipped and \$seconfi will be read next. If `jobtype` equals ‘bsopt’, then the \$newuoa group will be needed.

6. \$newuoa has to be given if `jobtype`=‘bsopt’. Refer to the newuoa [citation needed](#) documentation for detailed information if needed.

rhobeg	The initial value of the trust region used by newuoa (default=0.1).
rhoend	The final value of the trust region used by newuoa. Must be smaller than rhobeg (default= $1.0 \times 10^{-4}$ ).
iprint	The print level for newuoa.
= 0	No printing from newuoa (default).
= 1	Print only when newuoa has finished.
= 2	Print only when the trust region has decreased by an order of magnitude.
= 3	Print every iteration of newuoa.
maxfun	The maximum number of calls to calfun newuoa will make before terminating (default=500).

7. \$seconfi

nclose	The number of closed spinors for each symmetry.
nopen	The number of open spinors for each symmetry. There is a limit to one open orbital per symmetry.
freel	The number of electrons available in the open spinors.
autogen	Automatically generates all possible combinations of spinor occupancies (default=.false.).
nconf	The number of configurations to be read in (needed if autogen is false).

If autogen is false, then the next nconf lines will be the spinor occupancies. They will be given as real numbers with one configuration per line.

8. \$scf			
maxitr			The maximum number of SCF iterations (default=50).
ixtrp			The method of extrapolation.
	=	0	No extrapolation (default).
	=	1	Extrapolate the Fock matrix.
dfctr			Damping factor for Fock maxtrix (default=0.3).
thdll			Convergence limit for the large-large components of the density matrix (default= $1.0 \times 10^{-5}$ )
thdsl			Convergence limit for the small-large components of the density matrix (default= $1.0 \times 10^{-7}$ )
thdss			Convergence limit for the small-small components of the density matrix (default= $1.0 \times 10^{-9}$ )

# Chapter 3

## Basis Sets

### 3.1 Introduction

As stated in section 1.2.2, the use of Gaussian-type functions is the favored way of constructing basis sets for the use in molecular calculations. The use of basis sets introduces a new source of error, as we are now only approximating a wavefunction instead of directly numerically calculating it. One of the ways this error can manifest itself is in basis set superposition error (BSSE). BSSE appears when comparing the differences in energy between two different systems, as one might do to compare the difference in energy between reactants and product in a chemical reaction. Because we can not use a basis set with an infinite amount of basis functions, the basis set for the products and reactants in this example will be different (each will potentially have a different number of basis function centers, locations, etc.). This means that we are not comparing calculations with exactly the same level of accuracy, and the results may be invalidated because of this (see Section [ref FHF chapter here](#) for more details on BSSE and how to overcome it).

There is also the possibility of approximate linear dependancies in the basis set, which rears its head it head during the SCF process. This happens when the overlap between any two different functions is close to unity (at least as far as a double precision number is concerned) and will mean that there will be eigenvalues that are equal to zero when diagonalizing the Fock matrix. We cannot find the inverse of such a matrix, so therefore the functions leading to these eigenvalues of zero must get transformed out of the matrix. This is kind of bandaid solution though, as the resulting Fock matrix after this transformation could produce errors in the energy on the order of chemical accuracy. It can also lead to a very aggravating process during a geometry optimization.

For example, say we have a molecular system in geometry state A. In this state, there are no approximate linear dependancies, and the geometry optimizer determines that the system would prefer to be in geometry state B. It updates the geometry, but in state B, some atoms

might be too close together and now there are eigenfunctions that need to be rejected. The energy calculated for geometry state B is inaccurate, which causes the geometry optimizer to think that it should place the atoms back into state A. But now, everything is exactly the same as it was, and the geometry optimizer, with no memory at all of what just happened, thinks that the perfect geometry is state B again. There is no way out of this loop, and we will not be able to find a stationary state, and all because of a poor choice in basis set. It is therefore important for a computational chemist to be aware of these issues so they can identify them when they occur, but even more so for someone who wants to make a basis set, so that these issues can be avoided in the first place.

Designing a basis set that reduces the effects of BSSE is relatively straightforward, we can simply brute force the problem by adding more and more functions until they begin to cancel each other out (more on this in the discussion of this chapter). Designing one that avoids the issue of approximate linear dependencies however is more tricky. It is simply not possible to test every possible combination of atoms in every possible geometry to determine if a basis set has these problems. But there are some strategies we can use to avoid this problem.

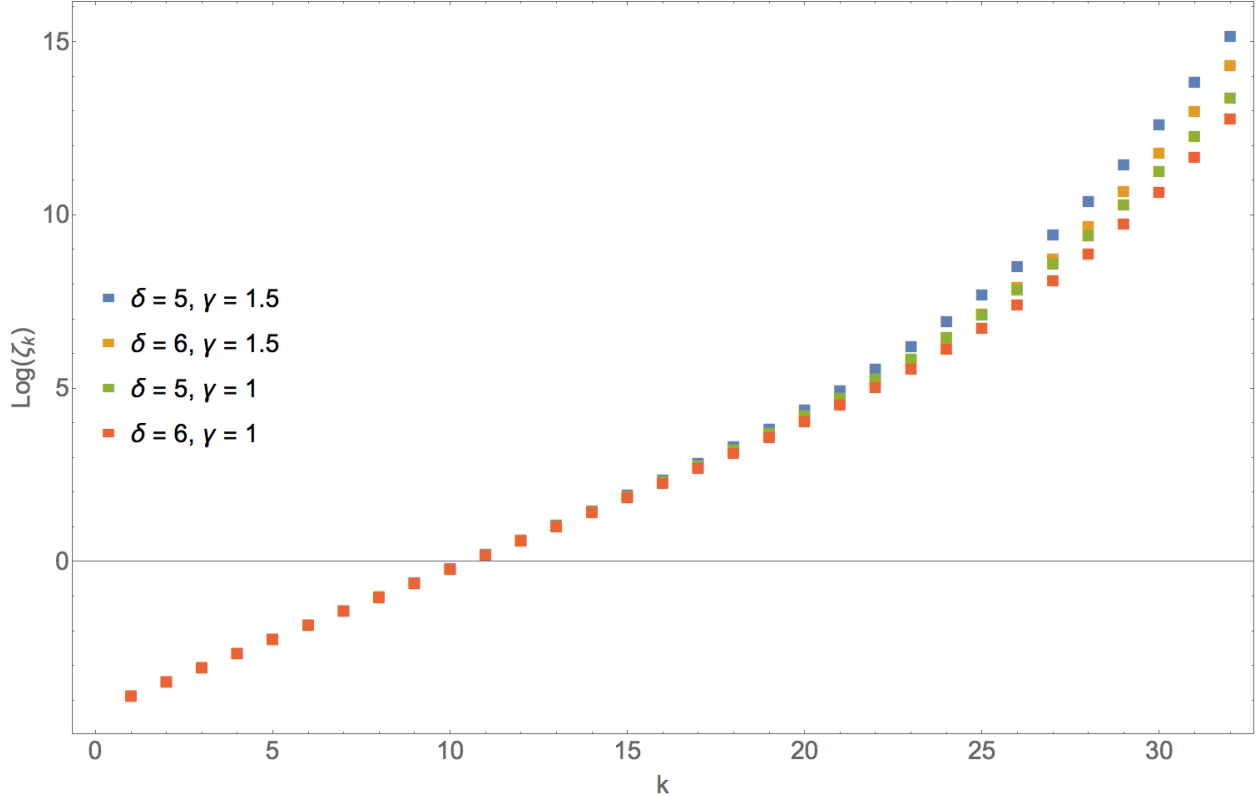
One strategy that has shown to be particularly effective is using an exponent generating function to ensure that the exponents of the basis set are sufficiently different. There are many of these generating functions<sup>[citation needed](#)</sup>. One of these give what are called even-tempered basis sets (ETBS). The generating function for an ETBS is given below

$$\zeta_{l,k} = \alpha_l \beta_l^{k-1}, \quad \alpha_l, \beta_l > 0, \quad \beta_l \neq 1, \quad k = 1, 2, \dots \quad (3.1)$$

where  $\zeta_{l,k}$  is the  $k^{\text{th}}$  exponent of basis function for orbital symmetry  $l$ ,  $\alpha_l$  is the exponent for the first, most diffuse, basis function, and  $\beta_l$  is a growth parameter that determines the gap between  $\zeta_{l,k}$  and  $\zeta_{l,k+1}$ . This greatly reduces the difficulty of the problem, because instead of needing to optimize all the exponents of our basis set, we only need to optimize the  $\alpha$  and  $\beta$  parameters of each symmetry. We can further reduce the number of parameters to optimize by using the same  $\alpha$  and  $\beta$  parameters *for* each symmetry. This is what I did for the work in this chapter, so I will drop the  $l$  subscript. ETBS has an issue though. The basis functions get exponentially more contracted the higher  $k$  becomes. So we get diminishing returns by increasing the maximum value of  $k$ , and we have to add many more functions to improve the accuracy of the basis set by even a small amount. This is improved in the well-tempered basis set (WTBS) generating function, which is shown below

$$\zeta_k = \zeta_{k-1} \beta \left[ 1 + \gamma \left( \frac{k}{n} \right)^\delta \right], \quad k = 2, 3, \dots, n, \quad \zeta_0 = \alpha, \quad \zeta_1 = \alpha \beta \quad (3.2)$$

where  $n$  is the maximum number of basis functions for the set, and  $\alpha$  and  $\beta$  have the



**Figure 3.1:** *Changes in  $\zeta$  as the values of  $\delta$  and  $\gamma$  are changed. The value of  $n$  was 36,  $\alpha$  was 0.02, and  $\beta$  was 1.5*

same restrictions on them as they did before. The parameters  $\delta$  and  $\gamma$  have the effect of shrinking the gap caused by  $\beta$  at high  $k$ . This means that the growth of the exponents is reduced as they get larger, which allows them to make a more meaningful contribution to the total wavefunction. Thus, we get a more accurate basis set with far fewer functions. The effect that altering  $\gamma$  and  $\delta$  has on the exponents can be seen in Figure 3.1.

One more thing to mention is that each symmetry does not need to use the entire set of  $\zeta$ s that we optimize. Indeed, the most diffuse orbitals in the valence shells often don't use the most contracted functions at all, and conversely the most contracted orbitals in the core don't make much use of the most diffuse functions. Therefore, it is better to think of the set of  $\zeta$ s as pool, and that each symmetry is drawing a range of functions from that pool.

## 3.2 Methods

### 3.2.1 Basis set size optimization

We decided to use a somewhat novel method of generating our basis sets. We begin by choosing an arbitrarily large basis set size, such as s(1 : 40)p(1 : 40)d(1 : 40)f(1 : 40) (here

$i(\# : \#)$  refers to the starting and ending index of  $\zeta$ 's used for symmetry  $i$ ). We then find the optimal  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  WTBS parameters for this basis set. We then use the following steps to find the optimal basis set size.

*Step 1.* Begin by finding the fewest number of f functions necessary. This can be done by generating input files that range in size from  $s(1 : 40)p(1 : 40)d(1 : 40)f(1 : 1)$  to  $s(1 : 40)p(1 : 40)d(1 : 40)f(1 : 40)$ .

*Step 2.* Optimize the basis sets and select the smallest basis set that is still below some minimum accuracy threshold. We chose a relative error of no greater than  $5.0 \times 10^{-9}$  to numerical calculations where relative error is defined as  $\frac{|(\text{optimized energy}) - (\text{numerical energy})|}{\text{numerical energy}}$ . The size of this basis set is  $s(1 : 40)p(1 : 40)d(1 : 40)f(1 : x_f)$ .

*Step 3.* Replace the WTBS parameters with those from the newly optimized set and generate a list of input files that range in size from  $s(1 : 40)p(1 : 40)d(1 : x_f)f(1 : x_f)$  to  $s(1 : 40)p(1 : 40)d(1 : 40)f(1 : x_f)$ .

*Step 4.* Optimize these new sets and select the smallest one that brings about an error that is still below the accuracy threshold. The size of this basis set is  $s(1 : 40)p(1 : 40)d(1 : x_d)f(1 : x_f)$ .

*Step 5.* Repeat steps 3 and 4 for the remaining symmetries. The size of the basis set at the end of this step will be of size  $s(1 : x_s)p(1 : x_p)d(1 : x_d)f(1 : x_f)$ .

*Step 6.* Replace the WTBS parameters with those from the newly optimized set and generate a list of input files that range in size from  $s(1 : x_s)p(1 : x_p)d(1 : x_d)f(1 : x_f)$  to  $s(1 : x_s)p(1 : x_p)d(1 : x_d)f(x_f : x_f)$ .

*Step 7.* Optimize and select from the basis sets. The new set will be of size  $s(1 : x_s)p(1 : x_p)d(1 : x_d)f(y_f : x_f)$ .

*Step 8.* Repeat steps 6 and 7 for the other symmetries except s. The final basis set will be of size  $s(1 : x_s)p(y_p : x_p)d(y_d : x_d)f(y_f : x_f)$ .

This procedure is perhaps a bit clunky, so let's work through an example using Ne. Ne has both s and p type orbitals, and so to will our basis set. We might initially guess that no more than 26 functions per symmetry will enough to get the accuracy we want, so we start with a basis set of size  $s(1:26)p(1:26)$ . We then optimize the  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  parameters for a basis set of this size. We make sure that the energy using a basis set of this size produces accurate enough results, and then we begin to trim off the fat. We carry over the WTBS parameters optimized in the last step, and generate input files corresponding to basis sets of size  $s(1:26)p(1:15)$  to  $s(1:26)p(1:26)$  (12 calculations in total) to try and determine the minimum number of contracted functions we will need to accurately describe the p orbitals. After reoptimizing the WTBS parameters, we find that while  $s(1:26)p(1:15)$  is too small to be accurate enough,  $s(1:26)p(1:16)$  is still inside the threshold we want. So, we take the WTBS parameters from this calculation, and use them to generate new basis sets ranging in

size from s(1:16)p(1:16) to s(1:26)p(1:16) (so we are trying to eliminate contracted functions from the s orbitals). Here we don't get as lucky as we did with the p orbitals, and we find that s(1:23)p(1:16) is the smallest basis set that we can use. Now, we want to eliminate diffuse functions. Carrying over the WTBS parameter from the last acceptable basis set, we generate basis sets of size s(1:23)p(1:16) to s(1:23)p(5:16). We find that trying to remove any more functions makes the basis set unusable so we stick with s(1:23)p(1:16). We could try to remove diffuse functions from the s symmetry, but experience teaches that s needs all the functions to be accurate, so we don't bother.

This process has a major drawback in that it requires a lot of computer power to run efficiently. But this is not really a problem if access to large computer clusters is available. Additionally, considering that calculations for the chain 1-8 can be run in parallel for different atoms, this optimization process is trivially parallel. The advantages of this process are twofold: it finds a very small basis set that is still accurate and it is completely automatable. If there are no problems with individual calculations not converging, the output files never even need to be manually (and intellectually) examined! We generated non-relativistic basis sets for elements 2-86 using the rwtbs program<sup>[citation needed](#)</sup> and relativistic basis sets for elements 55-86 using the cudaDFRATOM program which is outlined in detail in Chapter 2.

### 3.2.2 Automating Optimization

Automating the optimization process was quite simple. I did this by writing two scripts: BS\_sub\_auto\_rwtbs.sh and BS\_find\_best\_rwtbs.sh. The code for BS\_sub\_auto.sh is given here

```

1  #!/bin/bash
2
3  sym=''
4  alljobs=false
5  bottom=false
6  path_to_exe='/some/path' # must be set!
7  wtime=01:00:00
8
9  if [[ -z $path_to_exe ]]; then
10     echo "You must set the path to rwtbs.exe"
11     exit
12 fi
13
14 while [ $# -gt 0 ]
15 do
16     val=$1
17     shift
18     case $val in
19         "-a")
20             alljobs=true

```

```

21     ;;
22     "-b")
23         bottom=true
24     ;;
25     "-S")
26         sym='S'
27         low=$1
28         high=$2
29         shift
30         shift
31     ;;
32     "-P")
33         sym='P'
34         low=$1
35         high=$2
36         shift
37         shift
38     ;;
39     "-D")
40         sym='D'
41         low=$1
42         high=$2
43         shift
44         shift
45     ;;
46     "-F")
47         sym='F'
48         low=$1
49         high=$2
50         shift
51         shift
52     ;;
53     "-t")
54         wtime=$1
55         shift
56     ;;
57     *)
58         job=$val
59     ;;
60     esac
61 done
62
63
64 boilerplate(){
65     cat << EOF
66     #!/bin/bash
67     #SBATCH --account=def-mariusz # change depending

```



```

68 #SBATCH --mem=2GB
69 #SBATCH --time=$1
70 #SBATCH --job-name=$2
71
72 echo -----
73 echo -----
74 echo began running on \'date\'
75 echo -----
76 echo -----
77 $3/rwtbs $2 >& $4
78 echo -----
79 echo -----
80 echo stopped running on \'date\'
81 echo -----
82 echo -----
83 EOF
84 }
85
86
87 if [ "$alljobs" == true ]; then
88     job='ls *inp'
89 elif [ -z "$job" ]; then
90     echo no job given
91     exit
92 fi
93
94 for file in $job; do
95     dir=$(echo $file | sed -e "s/-.*//")
96     if [[ -d "$dir" ]]; then
97         echo "$dir already here"
98     else
99         mkdir $dir
100     fi
101     cd $dir
102
103     # keep old out files and delete others
104     rename "out" "out-old" *.out
105     rm *inp
106     rm *GUS
107     rm *MCS
108
109     cp ../$file $file"-orig"
110     file_base=$(echo $file | sed -e 's/_.*//')
111
112     # read the number of basis functions
113     read -r -a old_bs <<< $(sed -n -e '/nbfsym/{s/.*nbfsym(1)=//; s/\$end//; p}'
        $file"-orig")

```

```

114 # get the most contracted basis function
115 IFS='+' read -r -a old_top <<< $(sed -n -e '/ +/p' $file"-orig")
116
117 # keep track of how many files are made
118 total=0
119 for i in $(seq $low $high); do
120     for j in $(seq 0 3); do
121         new_bs[j]=${old_bs[j]}
122         new_top[j]=${old_top[j]}
123     done
124
125     # make alterations based on symmetry
126     case $sym in
127         "F")
128         ((new_bs[3]==$i))
129         # make sure that not too many functions are cut
130         if [[ "${new_bs[3]}" == "0" ]]; then
131             break
132         fi
133         ((new_top[3]+=$i))
134         ;;
135         "D")
136         ((new_bs[2]==$i))
137         if [[ "${new_bs[2]}" == "0" ]]; then
138             break
139         fi
140         ((new_top[2]+=$i))
141         ;;
142         "P")
143         ((new_bs[1]==$i))
144         if [[ "${new_bs[1]}" == "0" ]]; then
145             break
146         fi
147         ((new_top[1]+=$i))
148         ;;
149         "S")
150         ((new_bs[0]==$i))
151         if [[ "${new_bs[0]}" == "0" ]]; then
152             break
153         fi
154         ((new_top[0]+=$i))
155         ;;
156     esac
157
158     # make new calculation
159
160     new_file=$file_base_"${new_bs[0]}"s"${new_bs[1]}"p"${new_bs[2]}"d"${new_bs[3]}"f.inp"

```

```

160 cp $file"-orig" $new_file
161 sed -i "s/nalp=${old_bs[0]}/nalp=${new_bs[0]}/" $new_file
162 sed -i "s/${old_bs[0]} ${old_bs[1]} ${old_bs[2]} ${old_bs[3]}/${new_bs[0]}
    ${new_bs[1]} ${new_bs[2]} ${new_bs[3]}/" $new_file
163 if [ "$bottom" == false ]; then
164     sed -i -e "s/${old_top[0]}/${new_top[0]} +/" -e
    "s/${old_top[1]}/${new_top[1]} +/" -e "s/${old_top[2]}/${new_top[2]} +/" -e
    "s/${old_top[3]} ${new_top[3]} +/" $new_file
165 fi
166
167 # write submission script
168 boilerplate $wtime $new_file $path_to_exe ${new_file%???}out >
    ${new_file%???}.sh
169 #SBATCH ${new_file%???}.sh # uncomment to automatically submit job
170 ((total++))
171 done
172 echo "$total files made"
173 cd ../
174 done

```

**Listing 3.1:** *BS\_sub\_auto\_rwtbs.sh*

The first few lines handle the arguments given to the script. The arguments are flags which tell the script what to do. The flags can be given in any order, so long as each flag is followed by the appropriate options. The default flag is a file ending in “.inp” and is the calculation that you would like to run. Optionally, this flag can be omitted, which will tell the script to run over all “.inp” files in the current directory. I will refer to the input file to be used as a seed. The flags “-S”, “-P”, “-D”, and “-F” tell the script which set of orbitals we would like to remove functions from (“-S” means s orbitals, “-P” means p orbitals, and so on). These flags *must* be followed by two integers, with the first being smaller than the second. The first number is the minimum number for functions to cut, and the second is the maximum number to cut. By default, cutting begins with the most contracted function. For example, using “-S 0 3” would take the seed file with a basis set of size s(1:20) and produce otherwise identical files with basis sets of size s(1:20), s(1:19), s(1:18), and s(1:17). If the number are negative, this can be used to add basis functions (“-S -1 0” on the same seed would produce basis sets of size s(1:21) and s(1:20)). Using the “-B” (for bottom) flag changes the default behavior of cutting contracted functions, to cutting diffuse functions. So for example, using the flags “-S 0 3 -B” on the seed would produce s(1:20), s(2:20), s(3:20), and s(4:20).

Once the flags have been parsed, the script makes a directory with the same name as the title of the seed, moves inside this directory, and generates the needed input files. It also creates a job submission script using the boilerplate function for each of new input files and

submits them. Note that this part of the script will need to be changed depending on the cluster it is being used on. If more than one seed is used, then the script will do the same thing for each seed. After the jobs are completed, the user runs `BS_find_best_rwtbs.sh` in the directory that contains all of the newly created directories. The code for `BS_find_best_rwtbs.sh` is given below

```

1  #!/bin/bash
2
3  tol=$1
4
5  if [ -z $tol ]; then
6      tol='5*10^-9'
7  fi
8
9  echo The tolerance in relative error is $tol
10
11 mkdir best
12 for dir in `ls -d */`; do
13     if [ "$dir" != 'old/' ]; then
14         if [ "$dir" != 'best/' ]; then
15             if [ "$dir" != 'tests/' ]; then
16                 cd $dir
17                 for file in *.out; do
18                     error=`grep -A 1 '[E(WT)-E(rf)]/E(rf)' $file | tail -1 | awk '{print
19 $4}'`
20                     comperror=`echo $error | sed -e 's/D/\*10~/`
21                     testthress=$(bc -l <<< "scale=20; ($tol - $compperror)*10^10 >= 1")
22                     if [ $testthress -ge 1 ]; then
23                         cp `echo $file | sed -e 's/out/inp/'` ../best
24                         cp $file ../best
25                         echo Found best for $file
26                         break
27                     fi
28                 done
29             fi
30         fi
31     done
32 done

```

**Listing 3.2:** *BS\_find\_best\_rwtbs.sh*

This script takes in one argument, the relative error we are looking for in a calculation. It enters each of the subdirectories made by `BS_sub_auto.sh` and scans the output files for the relative error compared to numerical calculations. It finds the file with the smallest basis set that is still under the error threshold, and copies the input and output files for that calculation into a new directory called “best”. The user can then repeat this process for as

many orbitals as they wish. There are also similar scripts for cudaDFRATOM which are given in [Appendix whatever](#).

## 3.3 Discussion

### 3.3.1 Optimized Non-Relativistic Basis Sets

The optimized non-relativistic WTBS parameters and basis set sizes for elements 2-86 are shown in Table 3.1. Relativistic WTBS parameters for elements 55-86 are shown in Table 3.2.

**Table 3.1:** *Basis sets optimized using rwts. We did not have literature data for some elements. These elements are marked with a \*. The size of the basis set for these elements was determined by using the size of the nearest elements.*

Element	Term Symbol	$\alpha$	$\beta$	$\delta$	$\gamma$	s	p	d	f
02 He	$^1S$	$8.140 \times 10^{-2}$	1.953	4.504	1.515	(1:18)			
03 Li	$^2S$	$1.596 \times 10^{-2}$	1.933	5.701	1.573	(1:22)			
04 Be	$^1S$	$2.647 \times 10^{-2}$	1.938	5.841	1.594	(1:22)			
05 B	$^2P$	$3.238 \times 10^{-2}$	1.948	5.573	1.523	(1:22)	(1:15)		
06 C	$^3P$	$4.613 \times 10^{-2}$	1.941	4.717	1.317	(1:23)	(1:15)		
07 N	$^4S$	$5.976 \times 10^{-2}$	1.923	5.183	1.442	(1:23)	(1:16)		
08 O	$^3P$	$6.967 \times 10^{-2}$	1.936	5.170	1.426	(1:23)	(1:16)		
09 F	$^2P$	$8.321 \times 10^{-2}$	1.942	5.084	1.408	(1:23)	(1:16)		
10 Ne	$^1S$	$9.943 \times 10^{-2}$	1.945	4.988	1.392	(1:23)	(1:16)		
11 Na	$^2S$	$1.696 \times 10^{-2}$	1.950	6.056	1.469	(1:26)	(3:19)		
12 Mg	$^1S$	$2.375 \times 10^{-2}$	1.932	5.686	1.430	(1:26)	(3:19)		
13 Al	$^2P$	$2.239 \times 10^{-2}$	1.898	5.555	1.413	(1:27)	(1:20)		
14 Si	$^3P$	$3.352 \times 10^{-2}$	1.921	5.419	1.410	(1:26)	(1:19)		
15 P	$^4S$	$4.410 \times 10^{-2}$	1.907	5.196	1.390	(1:26)	(1:19)		
16 S	$^3P$	$5.004 \times 10^{-2}$	1.896	4.962	1.362	(1:26)	(1:19)		
17 Cl	$^2P$	$5.831 \times 10^{-2}$	1.886	4.784	1.344	(1:26)	(1:19)		
18 Ar	$^1S$	$6.834 \times 10^{-2}$	1.878	4.654	1.331	(1:26)	(1:19)		
19 K	$^2S$	$1.392 \times 10^{-2}$	1.889	6.354	1.523	(1:28)	(3:22)		
20 Ca	$^1S$	$1.837 \times 10^{-2}$	1.874	6.109	1.504	(1:28)	(3:22)		
21 Sc	$^2D$	$2.040 \times 10^{-2}$	1.878	6.160	1.505	(1:28)	(3:22)	(2:16)	
22 Ti	$^3F$	$2.218 \times 10^{-2}$	1.886	6.247	1.508	(1:28)	(3:22)	(2:16)	
23 V	$^4F$	$2.361 \times 10^{-2}$	1.887	6.573	1.507	(1:29)	(3:23)	(2:16)	
24 Cr	$^7S$	$2.345 \times 10^{-2}$	1.865	5.348	1.407	(1:29)	(3:22)	(2:17)	
25 Mn	$^6S$	$2.572 \times 10^{-2}$	1.865	5.336	1.407	(1:29)	(3:22)	(2:17)	
26 Fe	$^5D$	$2.621 \times 10^{-2}$	1.869	5.301	1.395	(1:29)	(3:22)	(3:17)	
27 Co	$^4F$	$2.643 \times 10^{-2}$	1.871	5.905	1.413	(1:29)	(4:23)	(3:17)	
28 Ni	$^3F$	$2.882 \times 10^{-2}$	1.875	5.935	1.413	(1:29)	(3:23)	(3:17)	

**Table 3.1:** (*continued*)

Element	Term Symbol	$\alpha$	$\beta$	$\delta$	$\gamma$	s	p	d	f
29 Cu	$^2S$	$2.283 \times 10^{-2}$	1.857	5.485	1.413	(1:30)	(4:23)	(3:18)	
30 Zn	$^1S$	$3.297 \times 10^{-2}$	1.881	5.817	1.339	(1:30)	(3:24)	(2:17)	
31 Ga	$^2P$	$2.742 \times 10^{-2}$	1.863	5.703	1.462	(1:30)	(1:23)	(3:18)	
32 Ge	$^3P$	$3.143 \times 10^{-2}$	1.848	5.290	1.386	(1:30)	(1:23)	(4:18)	
33 As	$^4S$	$4.728 \times 10^{-2}$	1.869	5.325	1.340	(1:30)	(1:23)	(3:17)	
34 Se	$^3P$	$5.087 \times 10^{-2}$	1.847	4.949	1.533	(1:30)	(1:22)	(3:18)	
35 Br	$^2P$	$5.927 \times 10^{-2}$	1.861	5.192	1.333	(1:30)	(1:23)	(3:17)	
36 Kr	$^1S$	$6.804 \times 10^{-2}$	1.859	5.510	1.370	(1:29)	(1:23)	(3:17)	
37 Rb	$^2S$	$1.421 \times 10^{-2}$	1.856	7.181	1.626	(1:30)	(3:25)	(6:21)	
38 Sr	$^1S$	$1.808 \times 10^{-2}$	1.846	6.674	1.553	(1:30)	(3:25)	(6:20)	
39 Y	$^2D$	$1.997 \times 10^{-2}$	1.841	6.569	1.543	(1:30)	(3:25)	(2:20)	
40 Zr	$^5F$	$2.184 \times 10^{-2}$	1.837	6.485	1.535	(1:30)	(3:25)	(2:20)	
41 Nb	$^6D$	$2.340 \times 10^{-2}$	1.835	6.434	1.530	(1:30)	(3:25)	(2:20)	
42 Mo	$^7S$	$2.537 \times 10^{-2}$	1.832	6.372	1.524	(1:30)	(3:25)	(2:20)	
43 Tc	$^6S$	$2.597 \times 10^{-2}$	1.836	6.453	1.534	(1:30)	(3:25)	(2:20)	
44 Ru	$^5F$	$2.682 \times 10^{-2}$	1.833	6.788	1.606	(1:30)	(3:25)	(2:21)	
45 Rh	$^4F$	$2.751 \times 10^{-2}$	1.835	6.883	1.620	(1:30)	(3:25)	(2:21)	
46 Pd	$^1S$	$5.944 \times 10^{-2}$	1.827	5.980	1.504	(1:29)	(2:24)	(1:19)	
47 Ag	$^2S$	$2.438 \times 10^{-2}$	1.800	5.521	1.423	(1:31)	(4:25)	(3:21)	
48 Cd	$^1S$	$2.911 \times 10^{-2}$	1.786	5.334	1.408	(1:31)	(4:25)	(3:21)	
49 In	$^2P$	$2.714 \times 10^{-2}$	1.799	5.540	1.431	(1:31)	(1:25)	(3:21)	
50 Sn	$^3P$	$2.884 \times 10^{-2}$	1.794	5.422	1.415	(1:31)	(1:25)	(4:21)	
51 Sb	$^4S$	$4.481 \times 10^{-2}$	1.815	6.567	1.605	(1:30)	(1:25)	(3:21)	
52 Te	$^3P$	$4.846 \times 10^{-2}$	1.817	6.247	1.528	(1:30)	(1:26)	(3:20)	
53 I	$^2P$	$5.369 \times 10^{-2}$	1.810	6.016	1.505	(1:30)	(1:25)	(3:20)	
54 Xe	$^1S$	$5.981 \times 10^{-2}$	1.802	5.862	1.490	(1:30)	(1:25)	(3:20)	
55 Cs	$^2S$	$1.177 \times 10^{-2}$	1.792	6.485	1.510	(1:33)	(4:28)	(5:23)	
56 Ba	$^1S$	$1.560 \times 10^{-2}$	1.770	6.152	1.518	(1:32)	(3:28)	(6:23)	
57 La	$^2D$	$1.716 \times 10^{-2}$	1.764	6.067	1.513	(1:32)	(3:28)	(2:23)	(5:14)
58 Ce*	$^3H$	$1.652 \times 10^{-2}$	1.774	6.254	1.529	(1:32)	(3:28)	(2:23)	(4:19)
59 Pr	$^4I$	$1.696 \times 10^{-2}$	1.776	6.297	1.533	(1:32)	(3:28)	(6:23)	(4:19)
60 Nd	$^5I$	$1.737 \times 10^{-2}$	1.778	6.563	1.579	(1:32)	(3:28)	(6:24)	(4:19)
61 Pm	$^6H$	$1.784 \times 10^{-2}$	1.780	6.615	1.585	(1:32)	(3:28)	(6:24)	(4:19)
62 Sm	$^7F$	$1.857 \times 10^{-2}$	1.778	5.842	1.464	(1:33)	(3:27)	(6:23)	(4:19)
63 Eu	$^8S$	$1.876 \times 10^{-2}$	1.784	6.718	1.596	(1:32)	(3:28)	(6:24)	(4:19)
64 Gd*	$^7F$	$1.926 \times 10^{-2}$	1.782	6.844	1.599	(1:33)	(3:28)	(6:24)	(4:19)
65 Tb*	$^6H$	$1.971 \times 10^{-2}$	1.784	6.898	1.604	(1:33)	(3:28)	(6:24)	(4:19)
66 Dy	$^5I$	$2.020 \times 10^{-2}$	1.788	6.543	1.540	(1:33)	(3:28)	(6:23)	(4:19)
67 Ho*	$^4I$	$2.064 \times 10^{-2}$	1.785	7.090	1.640	(1:33)	(3:28)	(6:23)	(4:20)
68 Er	$^3H$	$2.106 \times 10^{-2}$	1.792	7.029	1.638	(1:32)	(3:28)	(6:24)	(4:20)

**Table 3.1:** *(continued)*

Element	Term Symbol	$\alpha$	$\beta$	$\delta$	$\gamma$	s	p	d	f
69 Tm	$^2F$	$2.152 \times 10^{-2}$	1.794	7.075	1.643	(1:32)	(3:28)	(6:24)	(4:20)
70 Yb*	$^1S$	$2.200 \times 10^{-2}$	1.795	7.141	1.652	(1:32)	(3:28)	(2:24)	(4:20)
71 Lu	$^2D$	$2.373 \times 10^{-2}$	1.790	7.018	1.640	(1:32)	(3:28)	(2:24)	(4:20)
72 Hf	$^3F$	$2.511 \times 10^{-2}$	1.788	6.963	1.636	(1:32)	(3:28)	(2:24)	(5:20)
73 Ta	$^4F$	$2.684 \times 10^{-2}$	1.784	6.907	1.633	(1:32)	(2:28)	(2:24)	(5:20)
74 W	$^5D$	$2.814 \times 10^{-2}$	1.783	6.889	1.633	(1:32)	(3:28)	(2:24)	(5:20)
75 Re	$^6S$	$2.931 \times 10^{-2}$	1.782	6.886	1.634	(1:32)	(3:28)	(2:24)	(5:20)
76 Os	$^5D$	$3.075 \times 10^{-2}$	1.781	6.896	1.637	(1:32)	(2:29)	(2:24)	(4:20)
77 Ir	$^4F$	$3.195 \times 10^{-2}$	1.780	6.867	1.635	(1:32)	(3:28)	(2:24)	(5:20)
78 Pt	$^3D$	$3.173 \times 10^{-2}$	1.785	7.020	1.654	(1:32)	(3:28)	(2:24)	(5:20)
79 Au	$^2S$	$3.190 \times 10^{-2}$	1.791	6.702	1.569	(1:33)	(4:28)	(2:23)	(5:19)
80 Hg	$^1S$	$3.546 \times 10^{-2}$	1.781	6.691	1.604	(1:32)	(3:28)	(2:23)	(5:20)
81 Tl	$^2P$	$3.241 \times 10^{-2}$	1.794	6.826	1.561	(1:34)	(1:29)	(3:23)	(6:20)
82 Pb	$^3P$	$3.867 \times 10^{-2}$	1.778	6.948	1.656	(1:32)	(1:28)	(3:25)	(6:21)
83 Bi	$^4S$	$4.514 \times 10^{-2}$	1.766	6.158	1.545	(1:32)	(1:27)	(2:24)	(6:19)
84 Po	$^3P$	$4.783 \times 10^{-2}$	1.763	5.977	1.516	(1:32)	(1:27)	(3:23)	(6:19)
85 At	$^2P$	$5.210 \times 10^{-2}$	1.757	5.846	1.502	(1:32)	(1:27)	(3:23)	(6:19)
86 Rn	$^1S$	$5.716 \times 10^{-2}$	1.749	5.695	1.486	(1:32)	(1:27)	(3:23)	(6:19)

**Table 3.2:** *Basis sets optimized using cudaDFRATOM*

Element	Electron Configuration	$\alpha$	$\beta$	$\delta$	$\gamma$	s	p	d	f
55 Cs	[Xe] 6s <sup>1</sup>	$1.824 \times 10^{-2}$	1.921	6.085	1.062	(1:29)	(3:28)	(5:20)	
56 Ba	[Xe] 6s <sup>2</sup>	$2.033 \times 10^{-2}$	1.904	5.473	1.161	(1:30)	(3:28)	(6:20)	
57 La	[Ba] 4f <sup>1</sup>	$1.540 \times 10^{-2}$	1.871	5.166	2.036	(1:33)	(4:28)	(5:22)	(4:16)
58 Ce	[Ba] 4f <sup>2</sup>	$1.838 \times 10^{-2}$	1.857	4.969	1.450	(1:31)	(3:28)	(5:22)	(4:16)
59 Pr	[Ba] 4f <sup>3</sup>	$1.945 \times 10^{-2}$	1.857	4.947	1.438	(1:31)	(3:28)	(5:22)	(4:16)
60 Nd	[Ba] 4f <sup>4</sup>	$2.043 \times 10^{-2}$	1.859	4.928	1.422	(1:31)	(3:28)	(5:22)	(4:16)
61 Pm	[Ba] 4f <sup>5</sup>	$2.142 \times 10^{-2}$	1.860	4.906	1.405	(1:31)	(2:28)	(4:22)	(4:16)
62 Sm	[Ba] 4f <sup>6</sup>	$2.414 \times 10^{-2}$	1.856	4.854	1.379	(1:31)	(3:28)	(5:22)	(3:16)
63 Eu	[Ba] 4f <sup>7</sup>	$1.993 \times 10^{-2}$	1.877	5.099	1.020	(1:31)	(4:29)	(5:22)	(4:16)
64 Gd	[Ba] 4f <sup>8</sup>	$2.063 \times 10^{-2}$	1.880	5.064	0.996	(1:31)	(4:29)	(5:22)	(4:16)
65 Td	[Ba] 4f <sup>9</sup>	$2.396 \times 10^{-2}$	1.869	4.840	0.967	(1:31)	(3:29)	(5:22)	(4:16)
66 Dy	[Ba] 4f <sup>10</sup>	$2.177 \times 10^{-2}$	1.888	5.353	0.839	(1:30)	(4:29)	(5:22)	(4:16)
67 Ho	[Ba] 4f <sup>11</sup>	$2.517 \times 10^{-2}$	1.878	5.126	0.816	(1:30)	(3:29)	(5:22)	(4:16)
68 Er	[Ba] 4f <sup>12</sup>	$2.811 \times 10^{-2}$	1.875	5.145	1.045	(1:30)	(3:29)	(5:22)	(3:16)
69 Tm	[Ba] 4f <sup>13</sup>	$2.628 \times 10^{-2}$	1.887	5.480	0.886	(1:31)	(3:30)	(5:22)	(4:16)
70 Yb	[Ba] 4f <sup>14</sup>	$2.918 \times 10^{-2}$	1.883	5.490	0.873	(1:31)	(3:30)	(5:22)	(3:16)

**Table 3.2:** *(continued)*

Element	Electron Configuration	$\alpha$	$\beta$	$\delta$	$\gamma$	s	p	d	f
71 Lu	[Yb] 5d <sup>1</sup>	3.424×10 <sup>-2</sup>	1.873	5.124	1.353	(1:32)	(2:29)	(1:23)	(3:16)
72 Hf	[Yb] 5d <sup>2</sup>	3.982×10 <sup>-2</sup>	1.900	6.117	1.473	(1:32)	(2:29)	(1:22)	(4:16)
73 Ta	[Yb] 5d <sup>3</sup>	4.271×10 <sup>-2</sup>	1.899	6.126	1.450	(1:32)	(3:29)	(1:22)	(4:16)
74 W	[Yb] 5d <sup>4</sup>	4.564×10 <sup>-2</sup>	1.901	5.936	1.351	(1:32)	(3:29)	(1:21)	(4:16)
75 Re	[Yb] 5d <sup>5</sup>	4.827×10 <sup>-2</sup>	1.900	5.898	1.323	(1:32)	(3:29)	(1:21)	(4:16)
76 Os	[Yb] 5d <sup>6</sup>	5.081×10 <sup>-2</sup>	1.899	5.873	1.296	(1:32)	(3:29)	(1:21)	(4:16)
77 Ir	[Yb] 5d <sup>7</sup>	5.333×10 <sup>-2</sup>	1.898	5.846	1.274	(1:32)	(3:29)	(1:21)	(4:16)
78 Pt	[Yb] 5d <sup>8</sup>	5.391×10 <sup>-2</sup>	1.892	6.058	1.365	(1:32)	(3:29)	(1:23)	(5:16)
79 Au	[Yb] 5d <sup>9</sup>	5.684×10 <sup>-2</sup>	1.893	6.111	1.354	(1:32)	(3:29)	(1:23)	(5:16)
80 Hg	[Yb] 5d <sup>10</sup>	5.965×10 <sup>-2</sup>	1.893	6.156	1.342	(1:32)	(3:29)	(1:23)	(5:16)
81 Tl	[Hg] 6p <sup>1</sup>	4.071×10 <sup>-2</sup>	1.875	5.161	0.944	(1:30)	(1:29)	(1:22)	(5:17)
82 Pb	[Hg] 6p <sup>2</sup>	4.364×10 <sup>-2</sup>	1.869	5.015	0.925	(1:30)	(1:29)	(3:22)	(6:17)
83 Bi	[Hg] 6p <sup>3</sup>	5.109×10 <sup>-2</sup>	1.854	3.953	0.792	(1:33)	(1:30)	(2:21)	(5:17)
84 Po	[Hg] 6p <sup>4</sup>	5.439×10 <sup>-2</sup>	1.849	4.666	0.890	(1:30)	(1:29)	(3:22)	(6:17)
85 At	[Hg] 6p <sup>5</sup>	6.177×10 <sup>-2</sup>	1.892	10.223	1.500	(1:30)	(1:31)	(3:22)	(6:16)
86 Rn	[Hg] 6p <sup>6</sup>	6.671×10 <sup>-2</sup>	1.831	3.906	0.867	(1:31)	(1:29)	(3:21)	(6:17)

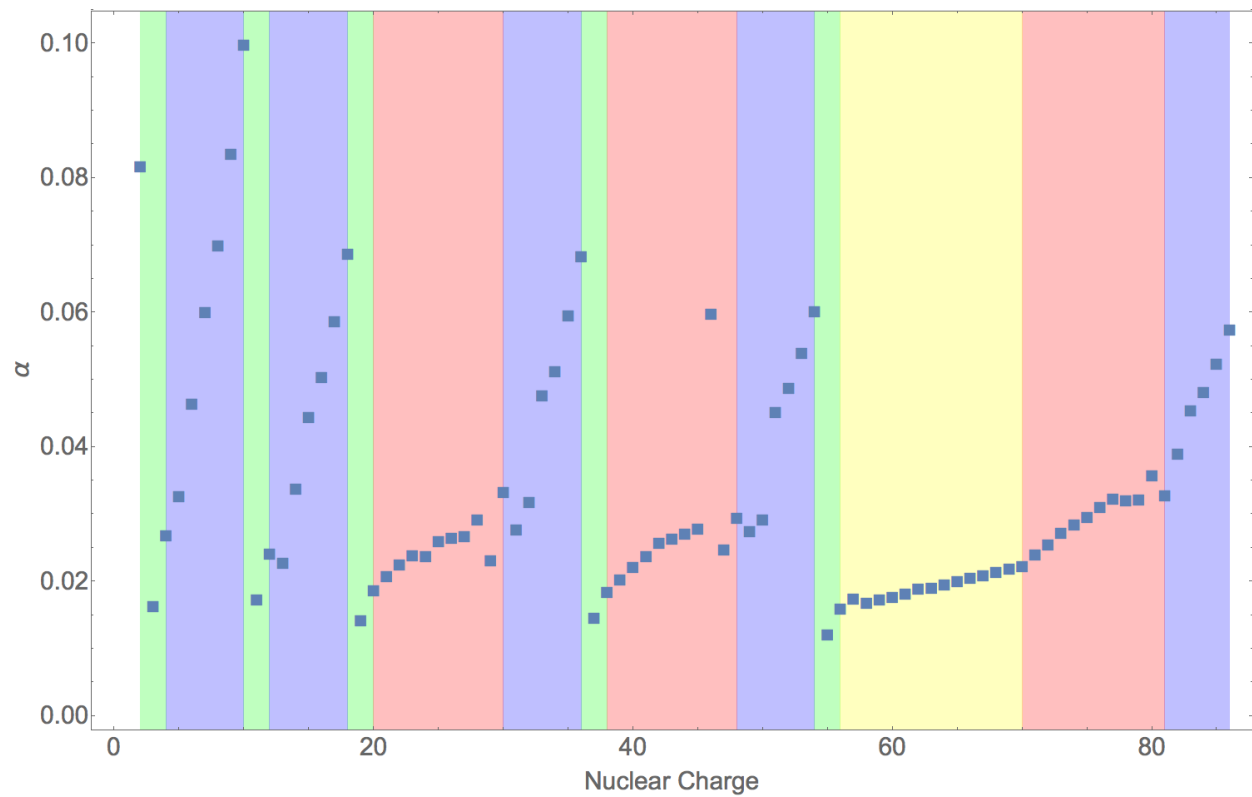
Figures 3.2, 3.3, and 3.4 show the trends the various non-relativistic parameters have with respect to nuclear charge. In Figure 3.2, we can see that there is an interesting pattern where the exponent of the most diffuse function becomes larger as we move right across a row. It is also interesting to note that the rate at which the exponent increases changes when going from one block of that row to the next. The trend roughly follows what we expect from periodic trends of atomic size. That is to say, that atoms get larger by going left on the table. The trend of atoms getting larger when moving down the periodic table is mostly obeyed, however there are exceptions. For instance, the 4d elements have smaller values of  $\alpha$  than the 5d elements.

In Figure 3.3 shows that  $\beta$  generally decreases as nuclear charge increases. As  $\beta$  is involved in controlling the growth of the exponents, this indicates that the basis sets are getting more diffuse as the nuclear charge increases.

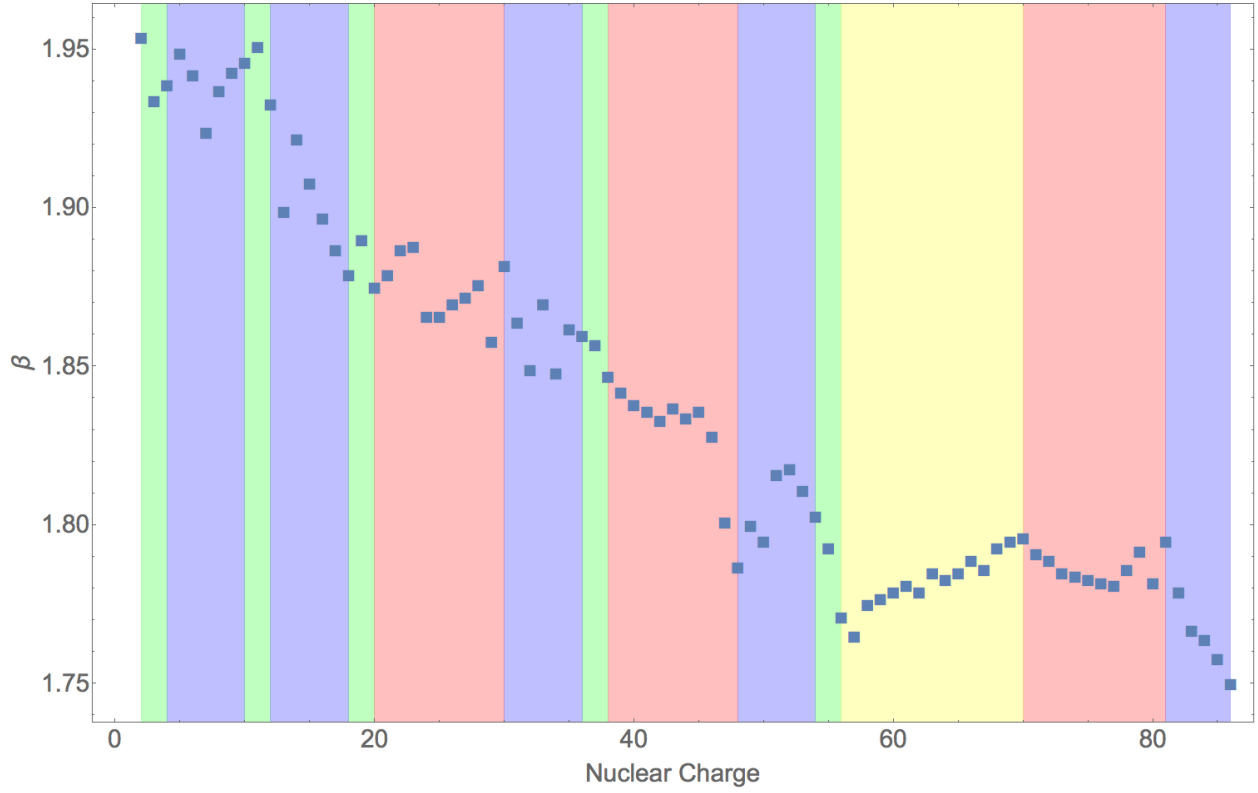
Finally 3.4 shows the values of  $\delta$  and  $\gamma$  with respect to nuclear charge. There are no clear trends in these parameters, but they are shown anyway for completeness.

As for the relativistic basis sets, we can see from Figure 3.5 that the trend in  $\alpha$  from the non-relativistic basis sets mostly holds, however there is a slight change in the trend starting at thallium. There is a substantial drop in the value of  $\alpha$  when changing from the d to the p block in the relativistic basis sets which is not seen in the non-relativistic sets. We can also see that the change in  $\alpha$  for the relativistic basis sets is much larger going

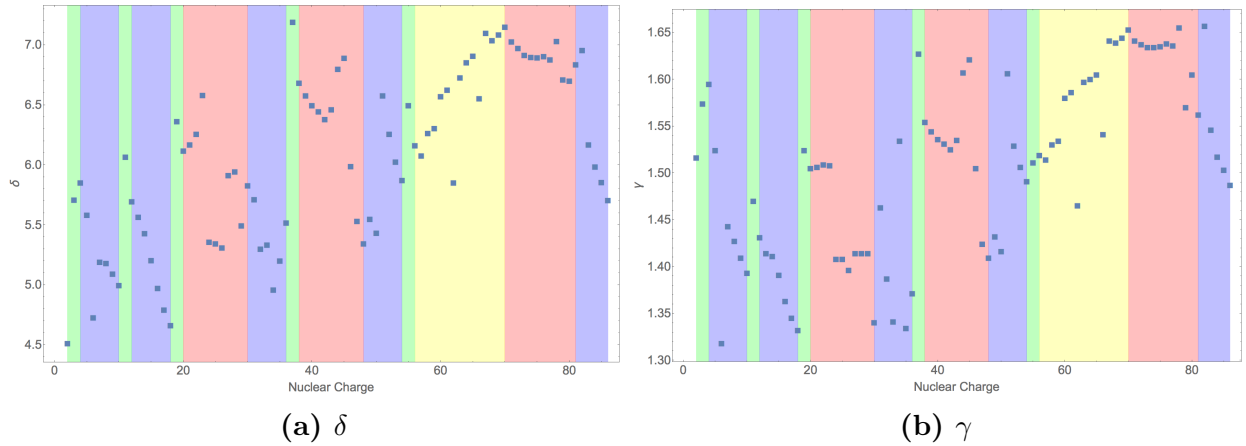




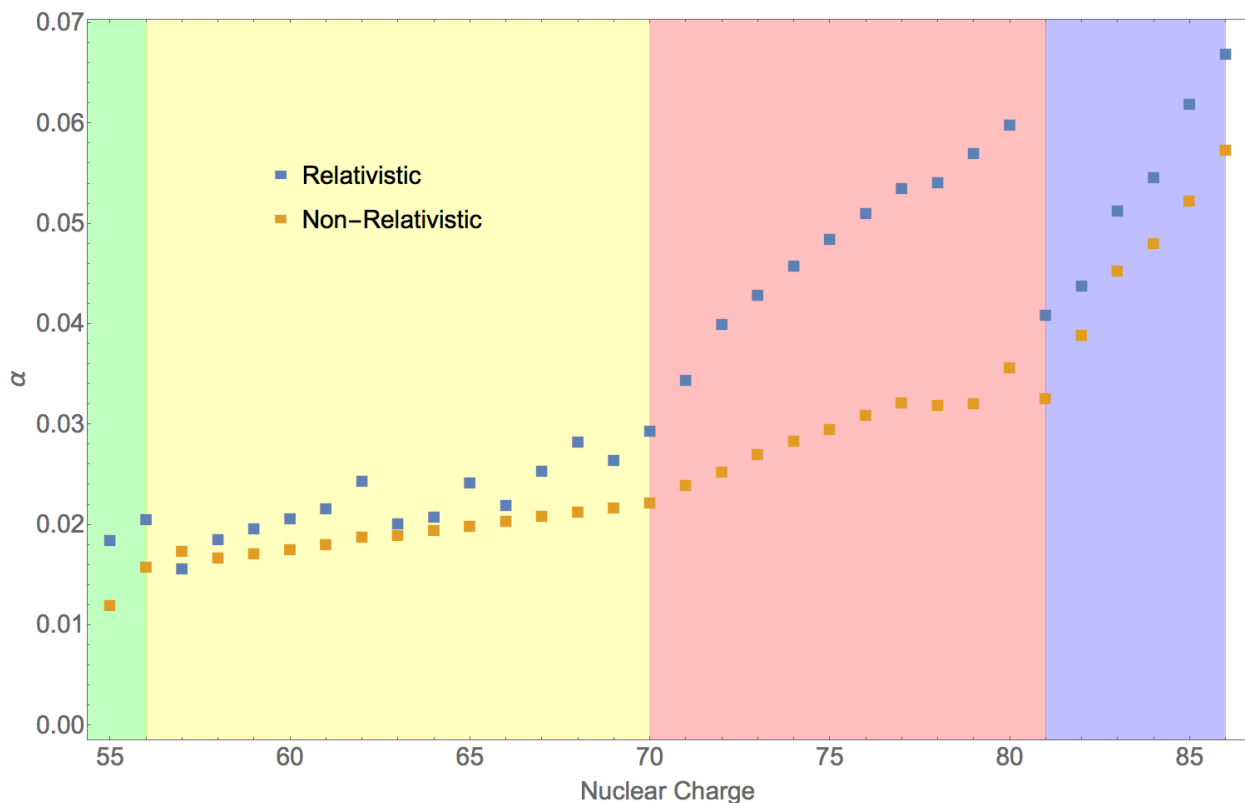
**Figure 3.2:** Scatter plot of the  $\alpha$  parameter vs. nuclear charge. Regions of the plot are shaded based on periodic table block regions. There is a distinct pattern where the value of  $\alpha$  increases along a row until a noble gas is reached. Also note how the rate of the increase changes at periodic table block boundaries.



**Figure 3.3:** Scatter plot of the  $\beta$  parameter vs. nuclear charge. Regions of the plot are shaded based on periodic table block regions. There is a general trend of the value of  $\beta$  to decrease as the nuclear charge grows larger.

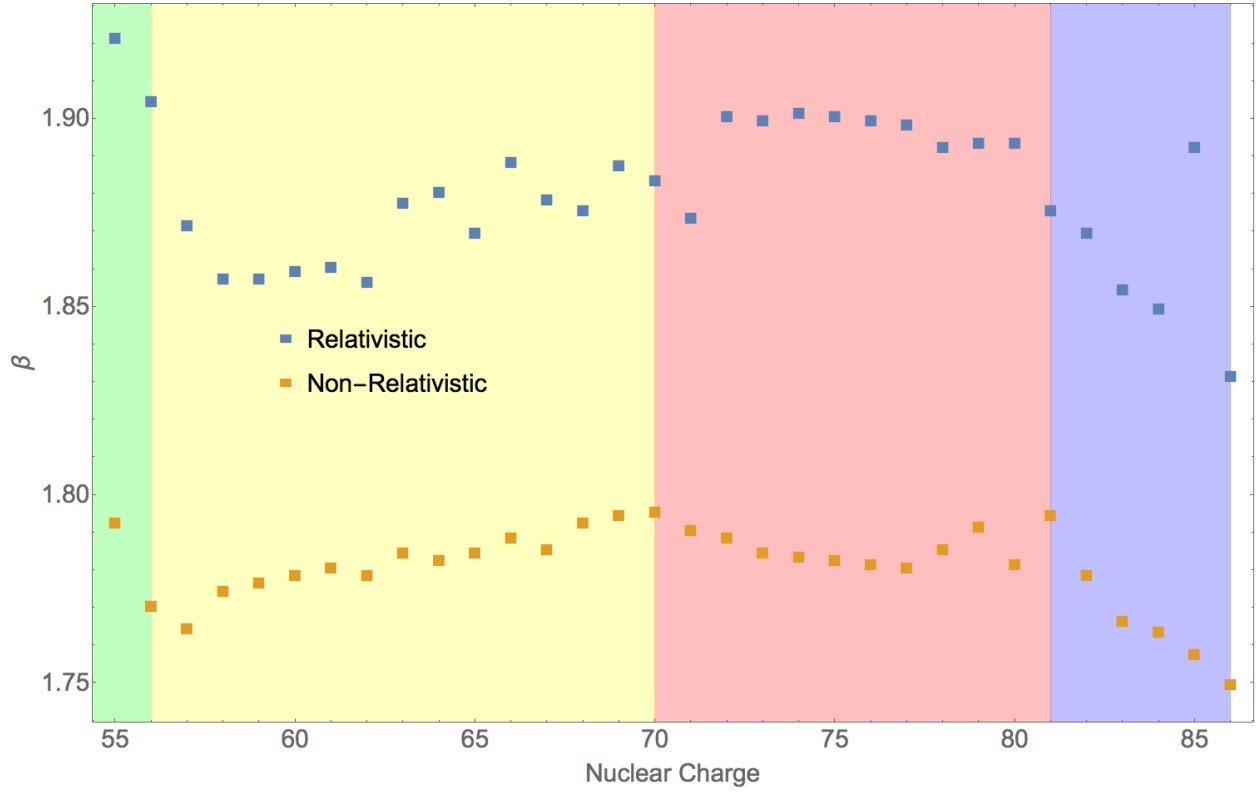


**Figure 3.4:** Scatter plots of the  $\delta$  and  $\gamma$  parameters. Regions of the plot are shaded based on periodic table block regions. There are no clear trends with respect to the nuclear charge.

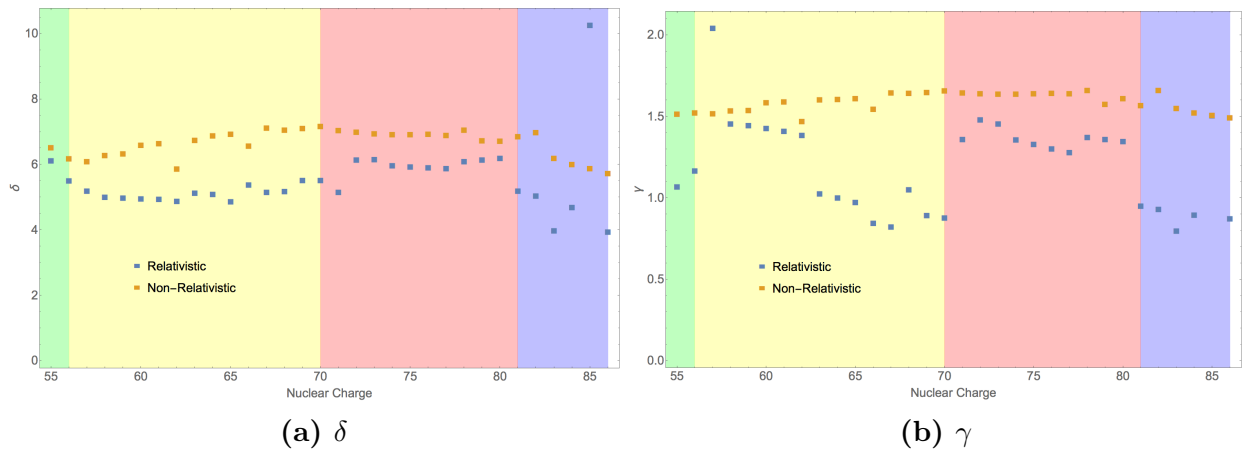


**Figure 3.5:** *Scatter plot of the relativistic and non-relativistic  $\alpha$  parameters vs. nuclear charge. Regions of the plot are shaded based on periodic table block regions.*

through the d block than the change for the non-relativistic basis sets. As for the other WTBS parameters, the trends in the relativistic basis sets are mostly consistent with their non-relativistic counterparts. Figures for the remaining parameters are given in Figures 3.6 and 3.7.



**Figure 3.6:** Scatter plot of the relativistic and non-relativistic  $\beta$  parameters vs. nuclear charge. Regions of the plot are shaded based on periodic table block regions.



**Figure 3.7:** Scatter plot of the relativistic and non-relativistic  $\delta$  and  $\gamma$  parameters. Regions of the plot are shaded based on periodic table block regions.

# Bibliography

- [1] E. Schrödinger. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys Rev*, 28:1049–1070, December 1926.
- [2] M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Physik*, 389:457–484, 1927.
- [3] D. J. Griffiths. *Introduction to Quantum Mechanics*. Pearson Prentice Hall, Upper Saddle River, NJ, 2nd edition, 2005.
- [4] S. Kais, D. R. Herschbach, N. C. Handy, C. W. Murray, and G. J. Laming. Density functionals and dimensional renormalization for an exactly solvable model. *J. Chem. Phys.*, 99(1):417–425, 1993.
- [5] J. Fourier. *Théorie analytique de la chaleur*. Chez Firmin Didot, père et fils, 1822.
- [6] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34:1293–1322, Nov 1929.
- [7] W. Gerlach and O. Stern. Der experimentelle nachweis der richtungsquantelung im magnetfeld. *Z. Physik*, 9(1):349–352, Dec 1922.
- [8] W. Pauli. Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren. *Z. Physik*, 31(1):765–783, Feb 1925.
- [9] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory*. Dover Publications, Mineola, N.Y, 1996.
- [10] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23:69–89, Apr 1951.
- [11] J. C. Slater. Atomic shielding constants. *Phys. Rev.*, 36:57–64, Jul 1930.
- [12] S. F. Boys. Electronic wave functions - i. a general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. Lond. A*, 200(1063):542–554, 1950.
- [13] G. G. Hall. The molecular orbital theory of chemical valency viii. a method of calculating ionization potentials. *Proc. R. Soc. Lond. A*, 205(1083):541–552, 1951.
- [14] A. A. Michelson and E. W. Morley. On the relative motion of the earth and of the luminiferous ether. *Am. J. Sci*, 34:333, 1887.
- [15] L. Piela. *Ideas of Quantum Chemistry*. Elsevier, Amsterdam, 2007.
- [16] W. Gordon. Der comptoneffekt nach der schrödingerschen theorie. *Z. Physik*, 40(1):117–133, Jan 1926.

- [17] O. Klein. Elektrodynamik und wellenmechanik vom standpunkt des korrespondenzprinzips. *Z. Physik A*, 41(6):407–442, Oct 1927.
- [18] P. A. M. Dirac. The quantum theory of the electron. *Proc. R. Soc. Lond. A*, 117(778):610–624, 1928.
- [19] C. D. Anderson. The positive electron. *Phys. Rev.*, 43:491–494, Mar 1933.
- [20] N. Bartlett. Relativistic effects and the chemistry of gold. *Gold Bull*, 31(1):22–25, Mar 1998.
- [21] L. J. Norrby. Why is mercury liquid? or, why do relativistic effects not get into chemistry textbooks? *J. Chem. Educ.*, 68(2):110, 1991.
- [22] L. Visscher and K. Dyall. Dirac-fock atomic electronic structure calculations using different nuclear charge distributions. *Atom. Data Nucl. Data Tabl.*, 67(2):207 – 224, 1997.
- [23] W. R. Johnson and G. Soff. The Lamb Shift in Hydrogen-like Atoms,  $1 \leq Z \leq 110$ . *Atom. Data Nucl. Data Tabl.*, 33:405, 1985.
- [24] G. Ruetsch and M. Fatica. *CUDA Fortran for Scientists and Engineers: Best Practices for Efficient CUDA Fortran Programming*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2013.
- [25] PGI. *CUDA Fortran Programming Guide and Reference*, Version 2018.
- [26] O. Matsuoka and Y. Watanabe. An atomic dirac-fock-roothaan program. *Comput. Phys. Commun*, 139(2):218 – 234, 2001.