



Benemérita Universidad Autónoma de Puebla

---

Facultad de Ciencias Físico Matemáticas

---

Entropía informacional de átomos: Implementación del  
método de integración numérica Las Vegas+

Tesis presentada al

**Colegio de Física**

como requisito parcial para la obtención del grado de

**LICENCIADO EN FÍSICA**

por

Santiago Alberto Flores Román

Asesorado por

Dr. Juan Manuel Solano Altamirano

Dr. Jorge Velázquez Castro

Puebla Pue.

Noviembre de 2021





Benemérita Universidad Autónoma de Puebla

---

Facultad de Ciencias Físico Matemáticas

---

Entropía informacional de átomos: Implementación del  
método de integración numérica Las Vegas+

Tesis presentada al

**Colegio de Física**

como requisito parcial para la obtención del grado de

**LICENCIADO EN FÍSICA**

por

Santiago Alberto Flores Román

Asesorado por

Dr. Juan Manuel Solano Altamirano

Dr. Jorge Velázquez Castro

Puebla Pue.

Noviembre de 2021



**Título:** Entropía informacional de átomos: Implementación del método de integración numérica Las Vegas+

**Estudiante:** SANTIAGO ALBERTO FLORES ROMÁN

COMITÉ

---

Dr. Javier Miguel Hernández López  
Presidente

---

Dr. Eduardo Moreno Barbosa  
Secretario

---

Dr. Julio Manuel Hernández Pérez  
Vocal

---

Dra. Beatriz Bonilla Capilla  
Suplente

---

Dr. Juan Manuel Solano Altamirano  
Dr. Jorge Velázquez Castro  
Asesores



*Mira la cumbre de la montaña, pero recuerda admirar las cosas bellas del camino  
sin olvidar los pasos que dejas atrás.  
Después de todo, el regalo es la experiencia y la cima es parte del viaje.*

*Santiago Alberto Flores Román*



# Dedicatoria

A mis abuelos Hebert y Socorro, a quienes siempre llevaré en mi corazón. Ellos me mostraron que una persona con voluntad es capaz de cambiar lo que desea en su vida. Antes de que descansen los dos juntos, les entrego este trabajo como muestra de lo que lograron marcar en mí, su sabiduría.

Para mi amada esposa Diana Marlén. Cada letra en este trabajo es un detalle que le entrego con mi corazón y cada idea expresada en el mismo es una muestra del futuro que deseo que tengamos juntos. Este es un primer paso, pero los siguientes serán nuestro vuelo.



# Agradecimientos

A mi asesor Dr. Juan Manuel Solano Altamirano, quien me guió y acompañó durante este proyecto y muchos otros anteriores a este. Con ellos, me ayudó a desarrollar las herramientas que serán las más útiles para el futuro.

A toda mi familia. Ellos me han apoyado constantemente durante toda la licenciatura y sé que no habría llegado hasta acá si no hubiera sido por ellos.

A mi amada esposa Diana Marlén. En los momentos de mayor tensión y presión, ella me ha ayudado mucho no solo durante el transcurso de la licenciatura, sino que también con el desarrollo de este trabajo. Incluso me enseñó conceptos que van más allá de la física, los químicos.



# Índice general

<b>Resumen</b>	<b>1</b>
<b>1. Introducción</b>	<b>3</b>
<b>2. Antecedentes</b>	<b>5</b>
2.1. Método Monte Carlo . . . . .	5
2.2. Método Las Vegas+ . . . . .	5
2.3. Entropía de Shannon . . . . .	9
2.3.1. Entropía relativa . . . . .	10
2.3.2. Aplicación de la entropía de Shannon en átomos y moléculas . . . . .	10
2.3.3. Suma de las entropías de Shannon . . . . .	12
2.4. Método Hartree-Fock . . . . .	12
2.5. Métodos post-Hartree-Fock . . . . .	15
2.5.1. Métodos perturbativos Møller-Plesset . . . . .	15
2.6. Funcionales híbridos . . . . .	16
2.6.1. Funcional B3LYP . . . . .	17
2.6.2. Funcional $\omega$ B97X-D . . . . .	17
2.7. Conjuntos base de gaussianas . . . . .	18
2.8. DensToolKit . . . . .	18
<b>3. Metodología</b>	<b>21</b>
3.1. Programa <i>dtkintegrate</i> . . . . .	21
3.1.1. Clase VegasIntegrator . . . . .	21
3.1.2. Función MAIN . . . . .	23
3.1.3. Definición de la diferencia de entropía informacional . . . . .	24
<b>4. Resultados</b>	<b>27</b>
<b>5. Conclusiones</b>	<b>31</b>

<b>A. Validación del programa <i>dtkintegrate</i></b>	<b>33</b>
A.1. Eficiencia de <i>dtkintegrate</i> . . . . .	33
A.2. Eficiencia de <i>dtkintegrate</i> contra el método Monte Carlo . . . . .	33
A.3. Entropías de Shannon en átomos utilizando <i>dtkintegrate</i> . . . . .	35
<b>Bibliografía</b>	<b>37</b>

# Resumen

En este trabajo se estudió la entropía de la información de Shannon en un conjunto de átomos en su estado base. Se buscó establecer la posible aplicación de la entropía de Shannon como una medida de la calidad de la descripción de la densidad electrónica. Para esto, se calcularon las densidades electrónicas de los átomos utilizando diversas bases y métodos de la química computacional. La entropía de Shannon se obtuvo a través de una integración numérica con el método Las Vegas+. Parte importante del trabajo consistió en la implementación eficiente de este algoritmo, además de adaptarlo a la suite *DensToolKit*, el cual está programado en C++ y es un código libre.

Los resultados obtenidos a raíz del cálculo de la entropía de Shannon de átomos y moléculas mostraron contradicciones con la propuesta de utilizarla como medida de la calidad de las bases, aunque también mostraron una aparente dependencia entre la entropía propuesta y las bases utilizadas.



# Capítulo 1

## Introducción

Dos de los conceptos medulares en la química computacional son la función de onda y la densidad electrónica, pues estas dos albergan toda la información electrónica del sistema, ya sea átomo o molécula, y el conocimiento de su comportamiento electrónico permite describir las causas y consecuencias de la formación y rompimiento de enlaces. Incluso, su importancia ha incitado a que varios autores hayan desarrollado programas o paquetes como *DensToolKit*, el cual se enfoca en calcular la densidad electrónica de un sistema específico y con ella obtener diversos funcionales derivados de la misma densidad, entre ellos, la entropía de Shannon de átomos y moléculas.

En la química computacional, la única manera hasta ahora conocida de aproximarse a la función de onda de un átomo o molécula es a través de métodos numéricos como Hartree-Fock (HF) o los métodos perturbativos Møller-Plesset, los cuales son MP2(full), MP3(full), MP4(SDQ), entre otros. La manera en que se utilizan los métodos para describir el comportamiento electrónico de átomos es primero componiendo una función de onda como un producto de orbitales moleculares conformadas por combinaciones lineales de orbitales atómicos y de modo que se respete el principio de exclusión de Pauli. Una manera de componer tales orbitales atómicos es a través de combinaciones lineales de funciones llamadas *funciones primitivas* y algunos ejemplos de estos son las bases 6-31G, 6-31+G\*\*, 6-31G(2df,p), aug-cc-pVDZ o aug-cc-pVTZ, a los cuales uno puede acceder a través de programas como Gaussian09 [1]. Estos son conjuntos de funciones de tipo gaussiano que al superponerse se aproximan a los orbitales atómicos. Es importante mencionar que estos conjuntos de funciones base deberán tener un número infinito de términos para conformar una base completa, por lo que cuantos más términos tengan, mejor podrá ser la descripción de los orbitales atómicos. Esto significa que la calidad una base dependerá del número de términos que la conformen y el uso específico (polarización, difusión, desdoblamiento) de estos términos para la descripción del orbital atómico, haciendo referencia a «calidad» como la consistencia entre las observaciones experimentales y la información del sistema obtenida por la función de onda calculada numéricamente.

Respecto a la densidad electrónica, su utilización surge por el desarrollo de la Teoría del Funcional de la Densidad (DFT por sus siglas en inglés), cuyo problema radica en el desconocimiento del funcional de la energía de correlación e intercambio ( $E_{CX}$ ) en la mayoría de los sistemas conocidos. No obstante, varios autores han sugerido aproximaciones a partir del acompañamiento de parámetros experimentales y de sistemas conocidos como un gas de electrones uniforme, donde este funcional sí se conoce y a volúmenes infinitesimales su comportamiento se asemeja a una mayor variedad de sistemas. Los funcionales B3LYP y  $\omega$ B97X-D presentan esas características y también son accesibles a través de Gaussian09. Respecto a la densidad electrónica, esta puede calcularse a través de los mismos conjuntos de funciones base mencionados en el párrafo anterior. Cabe mencionar que los funcionales en DFT habrán de depender esencialmente de la densidad electrónica.

Además del funcional de la energía de correlación e intercambio, existe una gran variedad de funcionales y uno de ellos es la entropía de Shannon, la cual también llamaremos en la presente como entropía informacional. Este objeto es de interés en la química computacional debido a varias razones, pero en la tesis presente, el enfoque se encuentra en que ha sido sugerida para medir la calidad de los conjuntos de funciones base. A pesar de que varios autores han estado estudiando esta utilidad [2–5], hasta ahora no se ha logrado concluir si es determinante para la medida de la calidad mencionada.

Por último, es importante comentar que la teoría DFT tiene la dificultad de que la densidad electrónica presenta desde uno a varios centros cuasi-singulares que complican su integración y hay funcionales que no son fáciles de conocer hasta el presente. Por consiguiente, no sería fácil utilizarlos como integrandos, tal como sucede con la entropía de Shannon. Entonces se vuelve necesario el uso de métodos de integración que no requieran previo conocimiento del comportamiento del integrando, como el método de integración Monte Carlo o sus derivados como Las Vegas+, método que se basa principalmente en el muestreo de importancia.

En la tesis presente se menciona la implementación del algoritmo de integración Las Vegas+ en el paquete de programas *DensToolKit* para probar si la entropía informacional es útil para medir la calidad de conjuntos de funciones base, de los cuales se utilizaron los mencionados al principio de este capítulo. Los sistemas de estudio fueron el conjunto de átomos del hidrógeno al kriptón y los métodos y funcionales utilizados también fueron los mencionados arriba. Primeramente, se desarrolló el método de integración Las Vegas+ y se implementó para cualquier función continua de tipo  $\mathbb{R}^3 \rightarrow \mathbb{R}$ , posteriormente se adaptó el mismo algoritmo al paquete de funciones *DensToolKit*, el cual permite calcular el integrando de la entropía de Shannon en los espacios de posiciones y momentos, y por último se analizó la entropía informacional de átomos del hidrógeno al kriptón para los métodos HF, MP2(full), MP3(full), MP4(SDQ), funcionales B3LYP,  $\omega$ B97X-D, así como los métodos 6-31G, 6-31+G\*\*, 6-31G(2df,p), cc-pVDZ, cc-pVTZ, aug-cc-pVDZ y aug-cc-pVTZ.

La estructura del trabajo presente se muestra de la siguiente manera. En el capítulo subsecuente se mencionan los conceptos requeridos para el desarrollo del proyecto, tales como los métodos de integración utilizados (Monte Carlo y Las Vegas+), la entropía de Shannon y sus variantes, donde se definen los conceptos y se presentan sus propiedades, los métodos HF y post-Hartree-Fock (Møller Plesset), así como los funcionales híbridos de la teoría DFT, los conjuntos de funciones base de tipo gaussiano y el paquete de programas *DensToolKit*. Una vez descritos los conceptos, se procede con la metodología del proyecto, especificando el desarrollo y la implementación del programa *dtkintegrate*, nombrado de esa forma por haberse adaptado al paquete *DensToolKit* y enfocarse en realizar integraciones sobre diferentes funcionales de la densidad electrónica. Se prosigue con el capítulo de resultados, donde se presentan los resultados obtenidos por parte del programa *dtkintegrate* y se realiza un análisis sobre los mismos y por último, el capítulo final aborda las conclusiones y propuestas para futuros proyectos relacionados con el objetivo principal de este proyecto. En la tesis presente se muestra un apéndice sobre los cálculos que se realizaron para conocer la eficiencia del método de integración Las Vegas+ y el programa *dtkintegrate* sobre densidades electrónicas.

## Capítulo 2

# Antecedentes

### 2.1. Método Monte Carlo

Dada una función  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , con  $d \in \mathbb{N}$ , se puede encontrar su integral sobre un volumen  $\Omega \in \mathbb{R}^d$  utilizando  $N_{ev}$  puntos aleatorios generados a partir de una densidad de probabilidad arbitraria  $p(\mathbf{x})$ , con  $\mathbf{x} \in \mathbb{R}^d$ . Numéricamente, es

$$\langle f^l \rangle = \frac{1}{N_{ev}} \sum_{N_{ev}} \left( \frac{f(\mathbf{x})}{p(\mathbf{x})} \right)^l, \quad (2.1)$$

siendo  $I$  la integral exacta de  $f$  sobre el volumen  $\Omega$ ,  $l \in \mathbb{N}$  y de modo que  $\langle f \rangle \rightarrow I$  cuando  $N_{ev} \rightarrow \infty$ .

La varianza para este método de integración se puede determinar a partir de la expresión

$$\sigma_{MC}^2 = \frac{\langle f^2 \rangle - \langle f \rangle^2}{N_{ev} - 1} \quad (2.2)$$

y consecuentemente, su desviación estándar es

$$\sigma_{MC} = \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N_{ev} - 1}}. \quad (2.3)$$

La ecuación (2.2) muestra la tendencia hacia la integral  $I$  conforme incrementan los puntos de muestreo que se realizan sobre  $\Omega$ . Esta tendencia puede resultar costosa en rendimiento computacional al ralentizarse a un ritmo de

$$\sqrt{\frac{1}{N_{ev}}}, \quad (2.4)$$

es por eso que se han desarrollado métodos adaptativos que buscan reducir ese costo sin sacrificar las ventajas del mismo método.

### 2.2. Método Las Vegas+

Este método de integración fue desarrollado por primera vez por Lepage, P. [6] en 1978 y ha sido tan ampliamente utilizado que incluso llegó a usarse en el CERN [7] y se implementó en diversos

lenguajes de programación como en Fortran [8] o Python a través del programa VegasFlow [9]. Además, se propusieron mejoras de la primera versión por parte de otros desarrolladores [10] para el área de la física de partículas. La última mejora implementada sobre este método fue publicada por su propio autor en agosto del 2021 y lo llamó Las Vegas+ [11]. La notación que se utilizará en esta sección será aquella utilizada por el autor del presente método, a excepción de la integral y varianza que dependen de las iteraciones posteriores, así como la varianza calculada por cada iteración.

La importancia de este método iterativo radica en su capacidad de adaptarse al integrando para reducir la incertidumbre  $\sigma^2$  prontamente, disminuyendo el número de evaluaciones a un orden  $O(N_g d)$ , en lugar de  $O(N_g^d)$ , donde  $N_g$  y  $d$  son el número de incrementos y de dimensiones, respectivamente.

Dada una una función  $f$  continua en un intervalo  $(a, b)$ , su integral puede obtenerse a través de un remapeo que la exprese en el intervalo  $(0, 1)$ ,

$$I = \int_a^b dx f(x) \Rightarrow T = \int_0^1 dy J(y) f(x(y)), \quad (2.5)$$

donde  $J$  es el jacobiano de la transformación. Para su aplicación en la computación, la segunda integral puede obtenerse a través de una aproximación discreta, es decir,

$$I \approx S = \frac{1}{N_{ev}} \sum_y J(y) f(x(y)), \quad (2.6)$$

donde  $N_{ev}$  es el número de evaluaciones que se realizan sobre  $f$ .

La integral  $I$  es el valor esperado de  $J(y)f(x(y))$ , por lo que la incertidumbre que presenta este método de integración es

$$\sigma_{MC}^2 = \frac{1}{N_{ev} - 1} \left( \sum_y [J(y)]^2 [f(x(y))]^2 - S^2 \right) \quad (2.7)$$

Se escoge tal transformación para minimizar la incertidumbre al integrar a través del método Monte Carlo y además se genera un mallado en el espacio original que varía en su forma (pero no en tamaño), de modo que

$$\begin{aligned} x_0 &= a \\ x_1 &= x_0 + \Delta x_0 \\ &\dots \\ x_{N_g} &= x_{N_g-1} + \Delta x_{N_g-1} = b. \end{aligned}$$

Tal mallado permite determinar la transformación, la cual es

$$x(y) = x_{i(y)} + \Delta x_{i(y)} \delta(y), \quad (2.8)$$

donde  $i(y)$  y  $\delta(y)$  son las partes entera y fraccionaria de  $yN_g$ , respectivamente,

$$i(y) \equiv \text{floor}(yN_g), \quad (2.9)$$

$$\delta(y) \equiv yN_g - i(y), \quad (2.10)$$

con  $y = i/N_g$  en el punto  $x_i$ . Así mismo, el jacobiano es

$$J(y) = N_g \Delta x_{i(y)}, \quad (2.11)$$

el cual es una función escalonada con anchos  $\Delta x_i$ .

En promedio, cada intervalo recibirá un mismo número de puntos de muestreo (alrededor de  $N_{ev}/N_g$ ), mostrando la aplicación del muestreo de importancia como método adaptativo en Monte Carlo.

El procedimiento del método Las Vegas+ consiste en generar un mallado uniforme y realizar un muestreo Monte Carlo para adquirir información del integrando, la cual será usada para modificar el mallado y realizar nuevamente un muestreo. Sucesivamente, el algoritmo recaudará cada vez más información para al final alcanzar un mallado óptimo y adaptado al comportamiento de la función a integrar.

La herramienta que extraerá información del integrando es

$$d_i \equiv \frac{1}{n_i} \sum_{x(y) \in \Delta x_i} [J(y)]^2 [f(x(y))]^2, \quad (2.12)$$

el cual deberá tender a una constante como índice de un mallado óptimo. Para realizar el refinamiento del mallado, Lepage, P. propuso que  $d_i$  se redefiniera como

$$d_i \rightarrow \left( \frac{1 - d_i}{\ln[1/d_i]} \right)^\alpha, \quad (2.13)$$

donde  $\alpha$  se define como tasa de convergencia y su valor depende del integrando, obteniéndose a prueba y error. Sin embargo, típicamente se escoge  $\alpha = 1$  y el valor  $\alpha = 0$  indica que ya no se realizará más refinamiento del mallado. Este parámetro es una constante durante el proceso de integración.

Sean

$$\begin{aligned} d &\equiv \sum_i d_i \\ &\text{y} \\ \delta d &\equiv \frac{\sum_i d_i}{N_g}. \end{aligned} \quad (2.14)$$

Una vez definidas las variables (2.14), se procede al siguiente algoritmo desarrollado por su autor.

1. Se inicializan las variables siguientes.

$$\begin{aligned} x'_0 &= x_0 \\ x'_{N_g} &= x_{N_g} \\ i &= 0 \text{ (índice del mallado } x') \\ j &= 0 \text{ (índice del mallado } x) \\ S_d &= 0 \text{ (Cantidad de } d \text{ acumulada)} \end{aligned}$$

2. Se incrementa  $i$  una unidad. Si  $i \geq N_g$ , el nuevo mallado ya terminó de generarse.
3. Se procede al siguiente paso si  $S_d \geq \delta d$ ; de otro modo, se agrega  $d_j$  a  $S_d$  y se incrementa  $j$  una unidad, volviendo al paso anterior.
4. Restamos  $\delta d$  a  $S_d$  y se calculan los límites del nuevo intervalo por interpolación:

$$x'_i = x_j - \frac{S_d}{d_{j-1}} \Delta x_{j-1}. \quad (2.15)$$

Se regresa al paso 2.

Una vez generada la nueva malla, se reemplaza la anterior por esta, se realiza un nuevo muestreo Monte Carlo y se repite el procedimiento hasta alcanzar el mallado óptimo y que el estimado de la integral sea suficientemente preciso.

Para funciones cuyo dominio se encuentre en un espacio multidimensional, es decir,

$$I = \int_0^1 d^D y J(y) f(\mathbf{x}(y)), \quad (2.16)$$

donde  $D$  representa el número de dimensiones que componen al espacio, los cambios que se hacen son

$$d_i^\mu \equiv \frac{1}{n_i^\mu} \sum_{\mathbf{x}^\mu(y^\mu) \in \Delta x_i^\mu} [J(y)]^2 [f(\mathbf{x}(y))]^2, \quad (2.17)$$

donde  $\mu$  representa la dimensión respectiva ( $\mu \in \{x, y, z\}$  para una función de tipo  $R^3 \rightarrow R$ ),  $0 < y^\mu < 1$  y

$$J(y) = \prod_{\mu} J^\mu(y). \quad (2.18)$$

Dado que cada integración Monte Carlo arroja información sobre el integrando, el autor del método Las Vegas propuso utilizar tal información para obtener un promedio y una varianza acumulados que ayudan a amortiguar la variación de ambas estimaciones, las cuales se pueden determinar mediante

$$I_{LV} = \frac{\sum_j I_j / \sigma_j^2}{\sum_j 1 / \sigma_j^2} \quad (2.19)$$

y

$$\sigma_{LV} = \left( \sum_j \frac{1}{\sigma_{MC,j}^2} \right)^{-1/2},$$

siendo  $j$  el índice del número de iteraciones.

Por último, se debe verificar que los resultados obtenidos en cada iteración sean consistentes con respecto a las incertidumbres estimadas. Lepage, P. propuso utilizar pruebas  $\chi^2$ , expresadas para este caso como

$$\chi^2 = \sum_j \frac{(I_j - \bar{I})^2}{\sigma_{MC,j}^2}. \quad (2.20)$$

La condición que su autor planteó fue que si  $\chi^2$  llegara a ser mayor al número de iteraciones presentes (menos uno), el proceso de integración deberá reiniciarse.

Hay casos donde  $\chi^2$  supera frecuentemente el umbral mencionado anteriormente. Algunas veces se debe a que la malla no se encuentra lo suficientemente refinada en las primeras iteraciones y otras veces a que el número de puntos de muestreo no es lo suficientemente grande para garantizar un comportamiento gaussiano alrededor de la integral. Para solucionar el primero, se puede eliminar la estimación obtenida en las primeras iteraciones (termalización) para considerar solo aquellas donde el mallado ya se encuentra lo suficientemente refinado; para el segundo caso, será suficiente con incrementar el número de puntos de muestreo a utilizar por iteración.

## 2.3. Entropía de Shannon

Este concepto, desarrollado por Shannon, C. E. [12] y nombrado como tal por su autor debido a su semejanza con la entropía definida en la mecánica estadística, se encarga de medir la cantidad de desinformación que presenta una variable aleatoria y matemáticamente se define como

$$H(p_1, p_2, \dots, p_n) \equiv -K \sum_{i=1}^n p_i \log p_i, \quad (2.21)$$

donde  $p_i(x) = P(X = x_i)$ , con  $X$  asociada al conjunto  $\mathbb{X}$ , es la probabilidad de ocurrencia de un evento  $i$ -ésimo,  $K$  es una constante positiva y se define  $p_i \log p_i \equiv 0$  cuando  $p_i(x) = 0$ . O bien, dado  $p(x) = P(X = x)$ , la entropía de Shannon también puede expresarse como

$$H(X) = -K \sum_{x \in \mathbb{X}} p(x) \log [p(x)]. \quad (2.22)$$

También existe la entropía para variables aleatorias conjuntas [13]. Dadas dos variables aleatorias  $X$  y  $Y$  con probabilidad conjunta  $p(x, y) = P(X = x, Y = y)$  y conjuntos  $\mathbb{X}$  y  $\mathbb{Y}$ , la entropía conjunta se define como

$$H(X, Y) \equiv -K \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} p(x, y) \log [p(x, y)], \quad (2.23)$$

mientras que para variables condicionales, la entropía condicional es

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathbb{X}} p(x) H(Y|X = x) \\ &= -K \sum_{x \in \mathbb{X}} p(x) \sum_{y \in \mathbb{Y}} p(y|x) \log [p(y|x)]. \end{aligned} \quad (2.24)$$

$H$  satisface las siguientes propiedades:

1.  $H$  es continua en cada probabilidad  $p_i$ , con  $i = 1, 2, \dots, n$ .
2. Si todas las probabilidades son iguales,  $p_i = 1/n$ , siendo  $n$  el número de eventos posibles, entonces  $H$  es una función monótona creciente y dependiente de  $n$ . Para probabilidades distintas,  $H \leq \log n$ . Esto quiere decir que los eventos con la misma probabilidad de ocurrir, tienen más incertidumbre cuando hay un mayor número de eventos posibles.
3.  $H(p_i) \geq 0$  cuando  $0 \leq p_i \leq 1$ . En caso de que  $H(p_i) = 0$ , el evento  $i$ -ésimo será seguro.
4. La entropía de Shannon también se puede interpretar como un valor esperado:

$$H(X) = K \langle \log (1/p(x)) \rangle_X, \quad (2.25)$$

donde el operador  $\langle \cdot \rangle_X$  hace referencia al valor esperado de la variable aleatoria  $X$ . Para variables conjuntas,

$$H(X, Y) = -K \langle \log [p(x, y)] \rangle_{XY}, \quad (2.26)$$

y para variables condicionales,

$$H(Y|X) = -K \langle \log [p(Y|X)] \rangle_{XY}. \quad (2.27)$$

5. La entropía de un par de variables aleatorias es la entropía de la una más la entropía condicional de la otra, es decir,

$$H(X, Y) = H(X) + H(X|Y). \quad (2.28)$$

6. La incertidumbre de dos variables conjuntas siempre será menor a la suma de las incertidumbres individuales, a menos que las segundas sean independientes.

$$H(X, Y) \leq H(X) + H(Y). \quad (2.29)$$

Esto quiere decir que la mínima incertidumbre que se puede alcanzar ocurre cuando los eventos son independientes entre sí. También se cumple que

$$H(X) \geq H(X|Y) \quad (2.30)$$

y nuevamente la igualdad se alcanza cuando ambas variables son independientes entre sí.

### 2.3.1. Entropía relativa

Más allá de las entropías mencionadas anteriormente, existe la entropía relativa o distancia *Kullback Leibler* entre dos funciones de masa de probabilidad  $p(x)$  y  $q(x)$  normalizadas a 1 [13], la cual se define como

$$H(p||q) \equiv \sum_{x \in X} p(x) \log \left[ \frac{p(x)}{q(x)} \right] = \langle \log(p/q) \rangle_X. \quad (2.31)$$

También se ha de definir  $H(p||q) \equiv 0$  cuando  $p = q$ , y cumple propiedades semejantes a su análoga absoluta, por ejemplo:

$$H(p||q) \geq 0 \quad (2.32)$$

y se alcanzará la igualdad si y solo si  $p(x) = q(x), \forall x \in X$ .

La entropía relativa  $H(p||q)$  representa la distancia entre las masas de probabilidad  $p$  y  $q$ , tomando a  $q$  como referencia.

Por último, las mismas entropías mencionadas en toda la sección tienen su equivalencia para densidades de probabilidad. Estas son

$$\begin{aligned} H(x) &= -K \int_{x \in X} p(x) \log [p(x)] dx, \\ H(x, y) &= -K \int_{x \in X} \int_{y \in Y} p(x, y) \log [p(x, y)] dy dx \\ &\quad \text{y} \\ H(p||q) &= \int_{x \in X} p(x) \log \left[ \frac{p(x)}{q(x)} \right] dx, \end{aligned} \quad (2.33)$$

para variables aleatorias continuas.

### 2.3.2. Aplicación de la entropía de Shannon en átomos y moléculas

La densidad electrónica de un átomo o molécula, comúnmente expresada como  $\rho(\mathbf{r})$  y definida como

$$\rho(\mathbf{r}) \equiv N_e \int d\sigma_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \psi^* (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \psi (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N), \quad (2.34)$$

donde  $\psi$  es la función de onda del sistema multielectrónico,  $\mathbf{x}_i = \mathbf{x}_i(\mathbf{r}_i, \sigma_i)$ ,  $\sigma_i$  es el espín del  $i$ -ésimo electrón,  $\mathbf{r}_i = \mathbf{r}_i(x, y, z)$  es una coordenada espacial arbitraria y  $N_e$  es el número de electrones que componen al átomo o la molécula, alberga la información del número de electrones por unidad de volumen en un punto del espacio de posiciones [14]. Cabe notar que se considera un espacio 4-dimensional (por cada electrón) sobre  $\psi$  al ser esta una función antisimétrica, por lo que el espín influye considerablemente en la localización de los electrones sobre el espacio 3-dimensional. Así mismo, la densidad electrónica puede estudiarse tanto en el espacio de posiciones, como en el espacio de momentos a través de la transformada de Fourier, y su expresión es  $\Pi = \Pi(\mathbf{p})$ . Esto debido a que el espacio es Fourier es el espacio de momentos:

$$\Pi(\mathbf{p}) = \frac{1}{(2\pi)^{3/2}} \int \rho(\mathbf{r}) e^{-i\mathbf{p}\cdot\mathbf{r}} d\mathbf{r}, \quad (2.35)$$

en unidades atómicas.

A sabiendas de la utilidad de la función (2.34), la entropía de Shannon puede otorgar información sobre el grado de localización de los electrones en un átomo o molécula: Cuanto mayor sea su valor, mayor incertidumbre habrá sobre sus posiciones o cantidades de movimiento. Conocida la densidad electrónica de un átomo o molécula, las entropías de Shannon en el espacio de posiciones  $S_\rho$  y en espacio de momentos  $S_\Pi$  pueden expresarse como [2]

$$\begin{aligned} S_\rho^{N_e}(\mathbf{r}) &= - \int \rho(\mathbf{r}) \ln [\rho(\mathbf{r})] d\mathbf{r} \\ &\text{y} \\ S_\Pi^{N_e}(\mathbf{p}) &= - \int \Pi(\mathbf{p}) \ln [\Pi(\mathbf{p})] d\mathbf{p}. \end{aligned} \quad (2.36)$$

Cabe remarcar que en ambas ecuaciones, las densidades no se encuentran normalizadas a 1. En tal caso, pasarán a ser [2]

$$\begin{aligned} S_\rho^1(\mathbf{r}) &= - \int \frac{\rho(\mathbf{r})}{N} \ln \left[ \frac{\rho(\mathbf{r})}{N} \right] d\mathbf{r} \\ &\text{y} \\ S_\Pi^1(\mathbf{p}) &= - \int \frac{\Pi(\mathbf{p})}{N} \ln \left[ \frac{\Pi(\mathbf{p})}{N} \right] d\mathbf{p}. \end{aligned} \quad (2.37)$$

También hay otras propuestas alternativas para estudiar la incertidumbre en las densidades electrónicas a partir de la entropía relativa, las cuales son para los espacios de posiciones y momentos, respectivamente,

$$\begin{aligned} S_{\rho_0}^{N_e}(\mathbf{r}) &= \int \rho(\mathbf{r}) \ln \left[ \frac{\rho(\mathbf{r})}{\rho_0(\mathbf{r})} \right] d\mathbf{r} \\ &\text{y} \\ S_{\Pi_0}^{N_e}(\mathbf{p}) &= \int \Pi(\mathbf{p}) \ln \left[ \frac{\Pi(\mathbf{p})}{\Pi_0(\mathbf{p})} \right] d\mathbf{p}, \end{aligned} \quad (2.38)$$

donde  $\rho_0$  y  $\Pi_0$  son densidades de referencia y por la definición de entropía relativa,  $S_{\rho,r}^{N_e}$  y  $S_{\Pi,r}^{N_e}$  miden la distancia entre ambas densidades correspondientes.

Otra propuesta, realizada por Flores Gallegos, N. [15], es usar como referencia el máximo de la densidad para el caso del estudio de un átomo o el promedio de los máximos locales para el

caso de moléculas. De ese modo, las densidades relativas  $\rho/\rho_{max}$  y  $\Pi/\Pi_{max}$  tomarían valores no negativos y menores a 1, refiriéndose al estudio de átomos. Por otro lado, aquellos máximos locales representan las densidades donde se ubican los núcleos, por lo que se medirían distancias entre la densidad en el núcleo y la densidad en el orbital de ocupación de mayor energía.

Las entropías pasarían a ser

$$S_{\rho_{max}}^1 \equiv -\frac{\rho_{max}}{N} \int \left[ \frac{\rho(\mathbf{r})}{\rho_{max}} \right] \log \left[ \frac{\rho(\mathbf{r})}{\rho_{max}} \right] d\mathbf{r}$$

y

$$S_{\Pi_{max}}^1 \equiv -\frac{\Pi_{max}}{N} \int \left[ \frac{\Pi(\mathbf{p})}{\Pi_{max}} \right] \log \left[ \frac{\Pi(\mathbf{p})}{\Pi_{max}} \right] d\mathbf{p}. \quad (2.39)$$

### 2.3.3. Suma de las entropías de Shannon

Białynicki-Birula, I. y Mycielski, J. [16] dedujeron que la suma de ambas entropías presentaba una equivalencia a la incertidumbre de Heisenberg, la cual es

$$S_{\rho}^{N_e} + S_{\Pi}^{N_e} \geq 3N_e (1 + \ln \pi) - 2N_e \ln N_e, \quad (2.40)$$

la cual muestra que habrá de presentarse una mínima incertidumbre entre las distribuciones de momentos y posiciones y depende del número de electrones del sistema. Sin embargo, si las densidades se encuentran normalizadas a 1, aquél mínimo alcanzable es

$$S_{\rho}^1 + S_{\Pi}^1 \geq 3(1 + \ln \pi). \quad (2.41)$$

Gadre, S. et al. [5] propusieron que la suma de las entropías podría arrojar información sobre la calidad de un conjunto base que conforma a los orbitales atómicos y puso a prueba su hipótesis con bases de desdoblamiento simple, doble y Near-Hartree-Fock [17]. Tripathi, A. et al. [4], realizaron una prueba semejante, pero con la serie isoelectrónica del Berilio. Por último, Hô, M. et al. [3] mostraron estudios sobre la sensibilidad de las entropías  $S_{\rho}$ ,  $S_{\Pi}$  y  $S_{\rho} + S_{\Pi}$  ante pequeñas variaciones de la densidad electrónica y sobre series isoelectrónicas de átomos.

## 2.4. Método Hartree-Fock

También llamado método del campo autoconsistente (SCF por sus siglas en inglés), consiste en encontrar la función de onda de un átomo a través de varias iteraciones [18–21]. En este método, se consideran funciones de onda antisimétricas y se expresan a través de determinantes de Slater.

Utilizando la notación de McQuarrie, D. A. [22], defínase la  $i$ -ésima función de onda espín orbital independiente del tiempo del  $j$ -ésimo electrón y con espín  $\alpha$  como  $\psi_i\alpha(j)$ , mientras que con espín  $\beta$ , es  $\psi_i\beta(j)$ . Además, sea  $\Psi(1, \dots, 2N)$  la función de onda de un átomo compuesto por  $2N$  electrones. Al dotarle la característica de antisimetría, esta puede expresarse como

$$\Psi(1, 2, \dots, 2N) = \frac{1}{\sqrt{(2N)!}} \begin{vmatrix} \psi_1\alpha(1) & \psi_1\beta(1) & \cdots & \psi_N\alpha(1) & \psi_N\beta(1) \\ \psi_1\alpha(2) & \psi_1\beta(2) & \cdots & \psi_N\alpha(2) & \psi_N\beta(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \psi_1\alpha(2N) & \psi_1\beta(2N) & \cdots & \psi_N\alpha(2N) & \psi_N\beta(2N) \end{vmatrix}, \quad (2.42)$$

el cual es el determinante de Slater, donde el factor a la izquierda del determinante es la constante de normalización de  $\Psi$  y  $\Psi$  se llamará función de onda determinantal.

Por otro lado, el operador hamiltoniano asociado a un átomo de  $2N$  electrones y carga nuclear  $Z$ , utilizando unidades atómicas, es

$$\begin{aligned}\hat{\mathcal{H}} &= -\frac{1}{2} \sum_{j=1}^{2N_e} \nabla_j^2 - \sum_{j=1}^{2N_e} \frac{Z}{r_j} + \sum_{j=1}^{2N_e} \sum_{j>i} \frac{1}{r_{ij}} \\ &= \sum_{j=1}^{2N_e} \hat{h}_j + \sum_{j=1}^{2N_e} \sum_{j>i} \frac{1}{r_{ij}},\end{aligned}\quad (2.43)$$

donde

$$\hat{h}_j \equiv -\frac{1}{2} \nabla_j^2 - \frac{Z}{r_j} \quad (2.44)$$

y

$$r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|. \quad (2.45)$$

Los primeros dos términos están asociados con la energía cinética de los electrones y la interacción electrón con núcleo, respectivamente, mientras que el tercero se debe a la interacción electrón con electrón.

El método Hartree-Fock resuelve cada espín orbital a través del operador de Fock  $\hat{F}$  (2.46). Al referirnos con  $\boldsymbol{\tau}_1$  y  $\boldsymbol{\tau}_2$  como las coordenadas espín-espaciales donde queremos ver la interacción entre dos electrones y al expresar  $\varphi_i$  como el  $i$ -ésimo espín orbital de un electrón,

$$\hat{F}(\boldsymbol{\tau}_1)\varphi_i(\boldsymbol{\tau}_1) = \varepsilon_i\varphi_i(\boldsymbol{\tau}_1), \quad i = 1, 2, \dots, N_e, \quad (2.46)$$

cuyo valor propio asociado al  $i$ -ésimo espín orbital es la energía orbital de Hartree-Fock,  $\varepsilon_i$ .

Si uno supone un sistema de capa cerrada, es decir que los orbitales se encuentran doblemente ocupados, el operador Fock puede depender únicamente de las coordenadas espaciales y expresarse como

$$\hat{F}(\mathbf{r}_1) = \hat{h}(\mathbf{r}_1) + \sum_j^{N_e} \left[ 2\hat{J}_j(\mathbf{r}_1) - \hat{K}_j(\mathbf{r}_1) \right], \quad (2.47)$$

donde los operadores de Coulomb  $\hat{J}$  y de intercambio  $\hat{K}$  son

$$\hat{J}_j(\mathbf{r}_1)\psi_i(\mathbf{r}_1) = \psi_i(\mathbf{r}_1) \int d\mathbf{r}_2 \psi_j^*(\mathbf{r}_2) \frac{1}{r_{12}} \psi_j(\mathbf{r}_2)$$

y

$$\hat{K}_j(\mathbf{r}_1)\psi_i(\mathbf{r}_1) = \psi_j(\mathbf{r}_1) \int d\mathbf{r}_2 \psi_j^*(\mathbf{r}_2) \frac{1}{r_{12}} \psi_i(\mathbf{r}_2) \quad (2.48)$$

y las funciones de onda orbitales ahora son espaciales,  $\psi$ . Estas son las *ecuaciones restringidas de Hartree-Fock*, sin embargo, no es necesario que se aplique la condición de capa cerrada para encontrar las soluciones, pues también se puede derivar el operador de Fock para casos de capa abierta y sus ecuaciones se llamarán *ecuaciones no restringidas de Hartree-Fock*.

A partir de la ecuación (2.46), se puede obtener la energía del  $i$ -ésimo orbital,

$$\varepsilon_i = \int d\mathbf{r}_1 \psi_i^*(\mathbf{r}_1) \hat{F}(\mathbf{r}_1) \psi_i(\mathbf{r}_1), \quad (2.49)$$

de la cual no se obtiene la energía total del sistema sumando todas ellas, sino a través de la ecuación (2.50),

$$E = \sum_{i=1}^{N_e} (I_i - \varepsilon_i), \quad (2.50)$$

donde

$$I_i = \int d\mathbf{r}_1 \psi_i^*(\mathbf{r}_1) \hat{h}_i(\mathbf{r}_1) \psi_i(\mathbf{r}_1) \quad (2.51)$$

deriva de la energía cinética del electrón y su interacción con el núcleo. Esto sucede porque en el método Hartree-Fock (HF) la interacción entre electrones se da a través de un potencial promedio, o bien, no hay correlación en el movimiento de los electrones, lo cual implica que no se considera la interacción instantánea entre electrones. Esta se calcula a través de la ecuación (2.52),

$$E_{corr} = E_{exacto} - E_{HF}, \quad (2.52)$$

la cual muestra la diferencia de energía entre la exacta y la obtenida usando el método HF. Cabe mencionar que la energía de correlación incrementa conforme incrementa el número de electrones que componen al sistema, alejando la energía calculada de la exacta.

Uno puede suponer que un orbital molecular se compone de una combinación lineal de orbitales atómicos (MO-LCAO por sus siglas en inglés),

$$\psi(\mathbf{r}) = \sum_i^k c_i \phi_i(\mathbf{r}), \quad (2.53)$$

además de que los orbitales atómicos  $\phi$  se conformarán de un conjunto de funciones base de tamaño arbitrario. Hay una gran variedad de estos conjuntos y cumplen diversas características propias [23].

Esta propuesta de función de onda implica trabajar con un sistema de ecuaciones llamadas *ecuaciones de Hartree-Fock-Roothaan* [21], las cuales pueden manipularse como matrices hermíticas de tamaño  $k \times k$ ,

$$\mathbb{F}\mathbf{c} = \varepsilon\mathbb{S}\mathbf{c}, \quad (2.54)$$

cuyos elementos son

$$\begin{aligned} \mathbb{F}_{ij} &= \int d\mathbf{r} \phi_i^* \hat{F}(\mathbf{r}) \phi_j(\mathbf{r}), \\ &\text{y} \\ \mathbb{S}_{ij} &= \int d\mathbf{r} \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}) \end{aligned} \quad (2.55)$$

y un vector columna  $\mathbf{c}$  de tamaño  $k$ . La matriz  $\mathbb{S}$  se llama *matriz de traslape*.

A través del método variacional, se encuentran los valores  $c_i$  que minimizan la energía del sistema y los valores obtenidos se utilizan para determinar las propiedades del sistema, es decir, la energía y los nuevos operadores de Fock. Tales propiedades se usan para buscar nuevos valores  $c_i$  que eventualmente con cada iteración, brinden una función de onda cada vez más consistente consigo misma, es decir, que la diferencia entre la energía de la iteración actual y la anterior se aproxime más a un valor nulo.

Cuanto más términos se utilicen en la ecuación (2.53), más se aproximarán los resultados a los exactos. Aunque no logren alcanzar valores exactos, estos tienden a un límite llamado *límite*

*Hartree-Fock*, en el cual se encontrarán las funciones de onda más precisas posibles que se pueden desarrollar utilizando este método.

La desventaja de este método es que considera una correlación promedio al suponer que un electrón interactúa con un campo electrostático promedio formado por el resto de electrones, lo cual ignora la dependencia entre los movimientos de los electrones con las posiciones del resto.

## 2.5. Métodos post-Hartree-Fock

Estos métodos se caracterizan por superar el límite Hartree-Fock al tomar en cuenta la correlación electrónica. Se llegan a utilizar diversas estrategias como excitaciones electrónicas o perturbaciones del estado Hartree-Fock, entre otras formas. Algunos de estos métodos son los siguientes.

### 2.5.1. Métodos perturbativos Møller-Plesset

Son métodos pos-Hartree-Fock que se caracterizan por usar la teoría perturbativa de Rayleigh-Schrödinger como estrategia para superar el límite Hartree-Fock y se pueden realizar perturbaciones de segundo orden (MP2) [24], tercero (MP3) [25], cuarto (MP4) [26] e incluso más. MP2 puede tomar en cuenta entre el 80 % y 90 % de la energía de correlación, MP3 entre el 90 % y 95 %, y MP4 entre el 95 % y 98 % [14]. Sobre los métodos subsecuentes, no se suele tener mucha información, puesto que requieren mucho trabajo computacional.

En estos métodos, el hamiltoniano no perturbado es la suma de los operadores de Fock provenientes del método HF,

$$\begin{aligned}\hat{\mathcal{H}}^{(0)} &= \sum_{j=1}^{N_e} \hat{F} \\ &= \sum_{i=1}^{N_e} \hat{h}_i + \sum_{i=1}^{N_e} \sum_{j=1}^{N_e} \langle \frac{1}{r_{ij}} \rangle,\end{aligned}\tag{2.56}$$

donde el último término se refiere a una interacción promedio. El hamiltoniano que se desea conocer es aquel de la ecuación (2.43) y el perturbado es

$$\begin{aligned}\hat{\mathcal{H}}' &= \hat{\mathcal{H}} - \hat{\mathcal{H}}^{(0)} \\ &= \left( \sum_{i=1}^{N_e} \hat{h}_i + \sum_{i=1}^{N_e} \sum_{j>i} \frac{1}{r_{ij}} \right) - \left( \sum_{i=1}^{N_e} \hat{h}_i + \sum_{j=1}^{N_e} \sum_{j=1}^{N_e} \langle \frac{1}{r_{ij}} \rangle \right) \\ &= \sum_{i=1}^{N_e} \sum_{j>i} \frac{1}{r_{ij}} - \sum_{i=1}^{N_e} \sum_{j>i} \langle \frac{1}{r_{ij}} \rangle.\end{aligned}\tag{2.57}$$

Entonces, por la teoría perturbativa de Rayleigh-Schrödinger, la energía de orden cero es

$$E_{HF} = E^{(0)} = \langle \psi_0 | \hat{\mathcal{H}}^{(0)} | \psi_0 \rangle,\tag{2.58}$$

la cual es la energía de Hartree-Fock y  $\psi_0$  el determinante de Hartree-Fock. La de primer orden es

$$E_{MP1} \equiv E^{(1)} = \langle \psi_0 | \hat{\mathcal{H}}' | \psi_0 \rangle,\tag{2.59}$$

la cual, por el teorema de Brillouin [27], no influye en una mejora con respecto al método Hartree-Fock, por lo que  $E_{MP1} = E_{HF}$ . La de segundo orden es

$$E_{MP2} \equiv E^{(2)} = 2 \sum_{i,j,a,b} \frac{\langle \varphi_i \varphi_j | \frac{1}{r_{12}} | \varphi_a \varphi_b \rangle \langle \varphi_a \varphi_b | \frac{1}{r_{12}} | \varphi_i \varphi_j \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} - \sum_{i,j,a,b} \frac{\langle \varphi_i \varphi_j | \frac{1}{r_{12}} | \varphi_a \varphi_b \rangle \langle \varphi_a \varphi_b | \frac{1}{r_{12}} | \varphi_j \varphi_i \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \quad (2.60)$$

donde  $\varphi_i$  y  $\varphi_j$  son orbitales ocupados,  $\varphi_a$  y  $\varphi_b$  orbitales virtuales (no ocupados en  $\psi_0$ ) y  $\epsilon_i$ ,  $\epsilon_j$ ,  $\epsilon_a$  y  $\epsilon_b$ , respectivamente, las energías orbitales asociadas a los orbitales apenas mencionados. Así sucesivamente, programas como Gaussian09 [1] alcanzan correcciones de orden mayor a cuatro.

Una observación a remarcar es que los determinantes de excitaciones dobles contribuyen para calcular la energía de segundo orden, los determinantes de excitaciones dobles y triples contribuyen para la de tercer orden, los de excitaciones dobles, triples y cuádruples contribuyen para la de cuarto orden, y así sucesivamente. Uno puede escoger arbitrariamente no considerar alguna de la excitaciones mencionadas en la perturbación de cuarto orden para reducir el costo computacional, en tal caso, se agregan las contribuciones que se tomaron en cuenta para el método MP4, por ejemplo, MP4(SDQ) ignora la contribución de determinantes de excitación triple (T) y solo toma en cuenta las simples (S), dobles (D) y cuádruples (Q).

Otra observación más es la reducción del costo computacional al cortar las sumas de la ecuación (2.60), lo cual también puede suceder para MP3, MP4 o subsecuentes. Cuando se consideran todos los términos de las sumas, estos métodos pueden renombrarse como MP2(full), MP3(full) o MP4(full), respectivamente.

## 2.6. Funcionales híbridos

En la teoría del funcional de la densidad (DFT por sus siglas en inglés), se establece que la densidad electrónica posee todas las propiedades electrónicas del sistema en estudio, de modo que se podría simplificar el estudio de átomos y moléculas al reemplazar a la función de onda, la cual depende de  $4n$  variables, con  $n$  como el número de electrones, por la densidad electrónica que depende de 3 variables.

Kohn, W. y Sham, L. [28] propusieron resolver el problema de una molécula a través de la consideración de un sistema de referencia  $s$  de infinitos electrones no interactuantes entre sí que ocuparan un volumen infinito y estuvieran sometidos a un potencial externo  $\nu_s$  (que sería causado por el núcleo), pero cuya densidad electrónica fuera aquella de la molécula en estudio en su estado fundamental,  $\rho$  que se puede conocer. Bajo estas circunstancias, la energía en función de la densidad sería

$$E[\rho] = \int \rho(\mathbf{r}) \nu_s(\mathbf{r}) d\mathbf{r} + T_s[\rho], \quad (2.61)$$

donde  $T_s[\rho]$  es el funcional energía cinética en función de  $\rho$ .

Para el caso real donde los electrones sí interactúan mutuamente bajo un potencial externo  $\nu$ , la energía sería

$$E[\rho] = \int \rho(\mathbf{r}) \nu(\mathbf{r}) d\mathbf{r} + T_s[\rho] + V_{ee}[\rho], \quad (2.62)$$

donde  $V_{ee}$  hace referencia a las interacciones electrón-electrón.

Comparando ambas ecuaciones, se obtiene la ecuación de la energía que se ha de encontrar para resolver el problema.

$$E[\rho] = T[\rho] + \int \rho(\mathbf{r})\nu(\mathbf{r})d\mathbf{r} + J[\rho] + E_{XC}[\rho], \quad (2.63)$$

donde

$$J[\rho] \equiv \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} \quad (2.64)$$

es la energía de repulsión electrostática y  $E_{XC}[\rho]$  es la energía de correlación e intercambio.

Para resolver el problema, se requiere conocer  $E_{XC}$ , sin embargo, hasta ahora sólo se saben aproximaciones de la misma a través de diversos métodos donde suponen que la energía  $E_{XC}$  es una combinación lineal de las energías de intercambio  $E_X$  y correlación  $E_C$  individuales y que además, ambas se deducen por diversos métodos, entre ellos, Hartree-Fock.

### 2.6.1. Funcional B3LYP

La energía funcional de B3LYP fue desarrollada por Stephens, P. et al. [29] y se compone de la combinación lineal

$$E_{XC}^{B3LYP} = (1-a)E_X^{LDA} + aE_X^{HF} + bE_X^{B88} + (1-c)E_C^{LDA} + cE_C^{LYP}, \quad (2.65)$$

donde  $E_X^{LDA}$ ,  $E_X^{HF}$  y  $E_X^{B88}$  son las energías de intercambio deducidas a través de los métodos de aproximación de la densidad local (*LDA* por sus siglas en inglés), Hartree-Fock (*HF*) y por Becke, A. (*B88*) [30], respectivamente, mientras que  $E_C^{LYP}$  [31] y  $E_C^{LDA}$  son sus respectivas energías de correlación. Los coeficientes  $a$  y  $b$  se ajustan con datos experimentales.

### 2.6.2. Funcional $\omega$ B97X-D

Este funcional fue desarrollado por Jeng-Da, C. y Head-Gordon, M. [32] para tomar en cuenta interacciones de dispersión (fuerzas de London) a través de correcciones de largo alcance y términos de amortiguamiento.

Para lograrlo, ajustaron el funcional  $\omega$ B97X [33], el cual contiene las correcciones de largo alcance, y un término asociado a la energía de dispersión, es decir,

$$E_{XC}^{\omega B97X-D} = E_{XC}^{\omega B97X} + E_{disp}, \quad (2.66)$$

donde la energía de dispersión es

$$E_{disp} = - \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{damp}(R_{ij}), \quad (2.67)$$

utilizando la notación de Jeng-Da, C. y Head-Gordon, M., donde  $N_{at}$  es el número de átomos que componen a la molécula,  $C_6^{ij}$  es el coeficiente de dispersión para la pareja de átomos  $ij$  y  $R_{ij}$  es la distancia interatómica.

La ecuación (2.67) contiene una función de amortiguamiento,

$$f_{damp}(R_{ij}) = \frac{1}{1 + a(R_{ij}/R_r)^{-12}}, \quad (2.68)$$

para tomar en cuenta los efectos de dispersión a corto alcance, donde  $R_r$  es la suma de los radios de van der Waals y  $a$  es un parámetro ajustable.

## 2.7. Conjuntos base de gaussianas

El conjunto base al que nos referiremos es a una combinación lineal de funciones que sirven para conformar a un orbital atómico. Existen varios tipos de conjuntos base, pero los más utilizados son los de tipo gaussiano (GTO) debido a la baja exigencia computacional que requieren, a diferencia de otros orbitales como los de Slater (STO) [34], cuyos cálculos son más exigentes para resolver las integrales (2.55), aunque describan mejor el comportamiento de  $\psi$ . Incluso, a pesar de que los orbitales GTO no representan adecuadamente el comportamiento electrónico de un sistema cerca del núcleo ni muy lejos del mismo, las combinaciones adecuadas de las mismas pueden arrojar un comportamiento aproximado a  $\psi$ .

Una función gaussiana primitiva en coordenadas cartesianas es

$$g_{ijk}(\mathbf{r}) = N x_b^i y_b^j z_b^k e^{-\zeta r^2}, \quad (2.69)$$

donde

$$N \equiv \left( \frac{2\zeta}{\pi} \right) \left[ \frac{(8\zeta)^{i+j+k} i! j! k!}{(2i)!(2j)!(2k)!} \right]^{1/2} \quad (2.70)$$

es la constante de normalización,  $x_b$ ,  $y_b$  y  $z_b$  son las coordenadas del núcleo  $b$ -ésimo,  $\zeta > 0$  es un parámetro ajustado con ayuda experimental y los valores  $i$ ,  $j$  y  $k$  indican el tipo de gaussiana que se desea crear (s, p, d o f). La razón de llamarla primitiva es porque se pueden formar orbitales atómicos GTO a partir de combinaciones lineales de funciones primitivas.

Las combinaciones lineales de estas funciones (2.69) se adecúan para cada orbital molecular e incluso llegan a reforzar a los orbitales de capas interna y de valencia formando combinaciones lineales de gaussianas para formar gaussianas más adecuadas para el caso, llamadas contracciones. La base más común es 6-31G [35], la cual utiliza una contracción de 6 gaussianas para los orbitales de capa interna, mientras que los orbitales de valencia se conforman por contracciones, una de 3 gaussianas primitivas en la zona interna y otra primitiva en la externa.

Para casos donde la nube electrónica se polariza, se agregan funciones orbitales de momento angular a los orbitales de valencia, como es el caso de la base 6-31G\* ó 6-31G(d), en el cual las funciones de polarización que se agregan son orbitales tipo d para los átomos pesados. Otro ejemplo es 6-31G\*\* ó 6-31G(d,p), donde se agrega además un orbital tipo p de polarización para los hidrógenos que compongan a la molécula. Por último, si hay presencia electrónica en zonas muy alejadas del núcleo, se agregan funciones de difusión que extiendan a la función de onda aún más, como la base 6-31+G, la cual agrega funciones de difusión tipo s y p para los átomos pesados. En el caso de la base 6-31++G, un función de difusión tipo s se agrega a los hidrógenos. Todas las bases GTO mencionadas anteriormente pueden combinarse entre sí para representar de manera más adecuada al sistema, por ejemplo 6-31+G\*\*.

Otras bases gaussianas conocidas son las de tipo cc-pVXZ [36], las cuales quieren decir correlación consistente con funciones X de valencia polarizadas, donde X puede ser doble (D), triple (T), cuádruple (Q), etcétera. Además, a estas bases también se les puede agregar gaussianas difusas, renombrándolas aug-cc-pVXZ. Estas bases se caracterizan por su uso en métodos que incluyen correlación electrónica más allá de lo permitido por el método HF, como son los métodos post-Hartree-Fock.

## 2.8. DensToolKit

*DensToolKit* es un paquete de programas con licencia GNU-GPLv3 y desarrollados por Solano-Altamirano, J. M. y Hernández-Pérez J. M. [37] enfocados en analizar densidades electrónicas y

varios campos derivados de ellos, tales como el gradiente de la densidad electrónica, la función de localización electrónica (ELF por sus siglas en inglés) junto con su gradiente, región de electrones lentos (RoSE por sus siglas en inglés), gradiente de densidad reducida, detector de electrones localizados (LED por sus siglas en inglés), entropías de la información, potencial electrostático, densidades  $K$  y  $G$  de energía cinética, entre otros, incluso en el espacio de momentos. Además, también incluye programas que se encargan de ubicar puntos críticos y enlaces químicos, todo en moléculas y átomos, usando como referencia la *Teoría Cuántica de Átomos en Moléculas* (QTAIM por sus siglas en inglés). Algunos de sus programas son

- `dtkpoint`, el cual evalúa los campos implementados en un punto o en un conjunto de puntos,  $y$
- `dtkbpdens`, cuya tarea es evaluar uno de los campos implementados a lo largo del trayecto del enlace entre dos átomos.

*DensToolKit* solicita como entrada archivos de formato *wfn* o *wfx*, los cuales son producidos por programas tales como Gamess, NWChem, MOLPRO o Gaussian. Tales archivos contienen la expansión de la función de onda molecular o atómica en términos bases gaussianas.

No obstante la variedad de sus programas, no hay uno que se encargue de integrar los campos escalares mencionados, por lo que es necesario implementar en él un algoritmo de integración para integrar cualquiera de los campos que ofrezca la suite.



# Capítulo 3

## Metodología

### 3.1. Programa *dtkintegrate*

Se desarrolló el algoritmo Las Vegas+ en lenguaje C++ y se implementó en DensToolKit como un programa llamado *dtkintegrate*, el cual llama a una clase nombrada *VegasIntegrator* que se encarga de aplicar el método de integración hacia el campo deseado y sobre el volumen especificado.

El algoritmo se compone de una clase llamada *VegasIntegrator* que se encarga de leer las instrucciones por parte del usuario y la función de onda a analizar, posteriormente utiliza tal información para integrar la densidad electrónica o el integrando de la entropía de Shannon. Cabe destacar que la función de onda las procesa el paquete DensToolKit para regresar la densidad electrónica en el espacio especificado por el usuario, así como la región de integración y otros campos que se deseen junto con sus puntos críticos, tales como las entropías informacionales y los puntos críticos de las densidades electrónicas. Es así que el algoritmo de integración recibe el campo requerido.

Los análisis realizados sobre la eficiencia del programa *dtkintegrate* se presentan en el apéndice de esta tesis.

#### 3.1.1. Clase *VegasIntegrator*

La clase da la disponibilidad al usuario de utilizar diversos métodos para una mayor personalización sobre el método de integración Las Vegas+.

Los métodos son los siguientes.

- Métodos que ingresan parámetros de entrada:
  - *SetDimensions*: Identifica la región de integración para funciones  $f$  del tipo  $f : R^3 \rightarrow R$ , la cual recibe de la información contenida en los archivos de entrada *wfx*, *wfn*.
  - *SetIntegrand*: Recibe al campo especificado como integrando. Por defecto, el campo es la densidad electrónica en el espacio de posiciones.
  - *SetIterations*: Si el proceso iterativo no alcanza el mallado óptimo, se detendrá al alcanzar un número de iteraciones arbitrario para presentar el estimado acumulado. Por lo general, se tienen 20 iteraciones, pues la integración alcanza una convergencia óptima antes de alcanzar las 20.
  - *SetIntervals*: Determina el tamaño fijo del mallado. Por defecto, son 10.

- **SetConvergenceRate:** Se establece la tasa de refinamiento de la malla. Este parámetro permite suavizar los cambios en el refinamiento del mallado y depende mucho del integrando, por lo que no se suele modificar. Por defecto, el valor es 1.
  - **SetNumOfPoints:** Determina el número de puntos de muestreo aleatorio Monte Carlo sobre la región de integración. Por defecto, son 10,000 evaluaciones por iteración.
  - **SetTermalization:** Indica el número de las primeras iteraciones cuya integral estimada no será utilizada para el estimado acumulado. Por defecto, su valor es 0.
  - **SetTolerance:** Dado que el mallado óptimo es no siempre se alcanza, se usa un intervalo de tolerancia. Por defecto, es 0.
  - **SetStopRefinement:** Dado que la estimación acumulada puede sesgarse, se puede cortar su cálculo a partir de cierta *j*-ésima iteración para basarse únicamente en los resultados de integración Monte Carlo, los cuales muestran una distribución gaussiana con media en la integral verdadera. Por defecto, no se toma un límite para este parámetro.
  - **SetNSamplesToFindMaximum:** Para buscar el valor máximo de la densidad en el espacio de momentos, se realiza un muestreo aleatorio sobre la región de integración con un número de puntos arbitrario e independiente de la integración Las Vegas+. Si se registra un valor de la densidad mayor a cero al lanzar el primer punto, se registra como máximo latente hasta que aparezca un valor registrado. Por cada valor registrado, el radio de búsqueda se reduce de la región de integración a la distancia entre el origen y el último valor registrado. La necesidad de realizar este método de rastreo se debe a que las densidades en el espacio de momento tienen comportamientos semejantes a una densidad gaussiana. El valor depende del átomo o la molécula en estudio, pues algunas veces el punto crítico se encuentra en el origen, lo que implica la innecesidad de solicitar este cálculo. En caso contrario, se suele requerir alrededor de 1,000,000 puntos de muestreo aleatorio para encontrar un valor aproximado.
- Método de visualización de parámetros de entrada y salida.
- **DisplayProperties:** Muestra todos los parámetros, tanto aquellos que se determinaron por el usuario como aquellos que se definen por defecto. Los parámetros son los límites de la región de integración, el integrando (el cual puede ser la densidad electrónica en el espacio de momentos o posiciones, la entropía de Shannon en alguno de ambos espacios, entre otros), la tasa de convergencia, tamaño de la malla, número de puntos de muestreo por iteración, máximo número de iteraciones, termalización, número de iteraciones para frenar el refinamiento de la malla y valor de la tolerancia para considerar la malla óptima.
  - **PrintLogFile:** Todos los datos de salida producidos por el programa *dtkintegrate* se imprimen en un archivo de extensión *log* y se guardan en donde se leyó el archivo de entrada *wfx/wfn*.
- Métodos que modifican a la densidad electrónica.
- **NormalizedEDF:** Normaliza la densidad electrónica a 1, al dividirla por el número de electrones. Esta constante de normalización la extrae el programa desde la suite *DensToolKit*.
  - **Relative2MaxDensity:** Calcula el valor máximo de la densidad. Si la densidad se encuentra en el espacio de posiciones, *dtkintegrate* utiliza información proporcionada por la suite *DensToolKit* para obtener los valores máximos; si se encuentra en el espacio de momentos, el mismo programa utiliza un proceso de rastreo del máximo a través de un muestreo aleatorio sobre la región de integración establecido en la función *SetSamplesToFindMaximum*.

- Métodos de salida de datos.
  - Variance: Muestra la varianza de la integral. El método de integración Las Vegas permite conocer tal varianza en cada una de las iteraciones y tal información ayuda a conocer la estabilidad en el proceso de integración a través de múltiples pruebas  $\chi^2$ . Pueden surgir casos donde la integración muestre signos de arrojar valores alejados de aquel esperado y uno de estos es la prueba  $\chi^2$ .
  - Integral: Muestra el resultado de la integral numérica. Dependiendo del tipo de integrando y si se desea que la densidad electrónica esté normalizada o sea relativa con respecto a su valor máximo, esta función se encarga de realizar transformaciones sobre la integral de la entropía o de la densidad.
  - MaxDensity: Muestra el valor máximo de la densidad electrónica. Si el cálculo es en el espacio de momentos, el valor es aquel obtenido por el método *Relative2MaxDensity*, en caso de haberse llamado. Si no, el valor máximo será el del origen o los núcleos, dependiendo del espacio.
  - NormConstant: Muestra la constante de normalización del integrando o de la densidad.
  - CountEvaluations y CountIterations: El primero muestra el número de evaluaciones hechas sobre el integrando antes de terminar el proceso, mientras que el segundo el número de iteraciones utilizadas para obtener el resultado.
- Método de ejecución.
  - Integrate: Esta función se encarga de realizar la integración de la función escogida sobre una región definida a través del método Las Vegas+. Se usa la herramienta generadora de números pseudo aleatorios Mersenne Twister [38], reconocida para algoritmos donde se usa gran cantidad de valores pseudo aleatorios; la información del integrando y la región de integración se obtienen a través de la clase *GaussWaveFunction*, perteneciente a la suite *DensToolKit*.

### 3.1.2. Función MAIN

El usuario puede agregar instrucciones a través de la línea de comandos en la terminal del sistema operativo (probado en GNU-Ubuntu 20.04 LTS, GNU-Debian 8 y MAC OSX v. 11 a 13), los cuales son la molécula o el átomo a estudiar (archivo *wfx* o *wfn*), la función que se desea integrar (densidades electrónicas en los espacios de posiciones y momentos, laplaciano de las densidades, entropías de Shannon en ambos espacios, energía cinética, potenciales electrostáticos, etcétera), el número de intervalos que compondrán la malla de integración, número de puntos a muestrear, número máximo de iteraciones, tasa de convergencia, termalización, tolerancia para considerar una malla óptima, número de iteraciones antes de detener el proceso de refinado de la malla, versión de *DensToolKit* e instructivo. Se habrá de notar que las funciones a integrar son campos procesados por *DensToolKit* y pueden utilizarse como integrandos, mas no necesariamente debe haber razón para su uso.

Además de la línea de comandos, aún el programa puede personalizarse más dentro de la función MAIN. De ahí, el usuario puede solicitar que se devuelvan la integral calculada (Integral), constante de normalización (NormConstant), densidad máxima (MaxDensity), varianza (Variance), número de iteraciones realizadas (CountIterations), número de evaluaciones hechas sobre el integrando (CountEvaluations) y todos los parámetros de entrada que se tienen por defecto, se calculan automáticamente o fueron ingresados por el usuario (DisplayProperties), así como ajustar la región de integración (SetDimensions), el cual, por lo general se calcula automáticamente.

Como ejemplo, la figura 3.1 muestra el cálculo de la integral de la densidad electrónica para el benceno, el cual tiene 42 electrones en su estado neutro. Dado que el archivo *benzene.wfn* es

utilizado por *DensToolKit* para probar la eficiencia de sus programas, *dtkintegrate* tiene registrado su valor exacto del número de electrones, por lo que puede calcular el porcentaje de error relativo entre su cálculo y aquel establecido como correcto.

### 3.1.3. Definición de la diferencia de entropía informacional

Para analizar los resultados, se tomará como hipótesis que la suma  $S_\rho^{N_e} + S_\Pi^{N_e}$  funge como medida de la calidad de las bases. Por otro lado, se definió la diferencia de entropías, donde se toma como referencia aquellas asociadas a la base 6-31G,

$$\Delta(S_T^{N_e})_i^{6-31G} \equiv (S_T^{N_e})_i^\mu - (S_T^{N_e})_i^{6-31G}, \quad (3.1)$$

donde  $(S_T^{N_e})_i^\mu$  es la suma de las entropías informacionales en los espacios de posiciones y momentos, asociada a la base  $\mu \in \{\text{aug-cc-pVDZ, aug-cc-pVTZ, cc-pVDZ, cc-pVTZ, 6-31G, 6-31+G^{**}, 6-31G(2df,p)}\}$  para el átomo  $i$ -ésimo,  $i \in \{1, 2, \dots, 36\}$ ,  $N_e$  indica la densidad electrónica normalizada al número de electrones del sistema, mientras que  $T$  indica la suma de ambas.

Se tomó como referencia 6-31G al ser la base que considera menos propiedades electrónicas a comparación de las bases restantes (polarización, funciones difusas, consistencia con la correlación electrónica e interacciones de largo alcance), por lo que valores negativos de  $\Delta(S_T^{N_e})_i^{6-31G}$  indicarían una calidad menor a aquella de la base 6-31G. Al comparar diferentes bases, aquella que muestre el valor  $\Delta(S_T^{N_e})_i^{6-31G}$  máximo para un átomo específico sería la base con la mejor calidad para tal átomo. Por último, la cantidad  $\Delta(S_T^{N_e})_i^{6-31G}$  para la base 6-31G valdrá cero para cualquier átomo o molécula por ser la base de referencia.





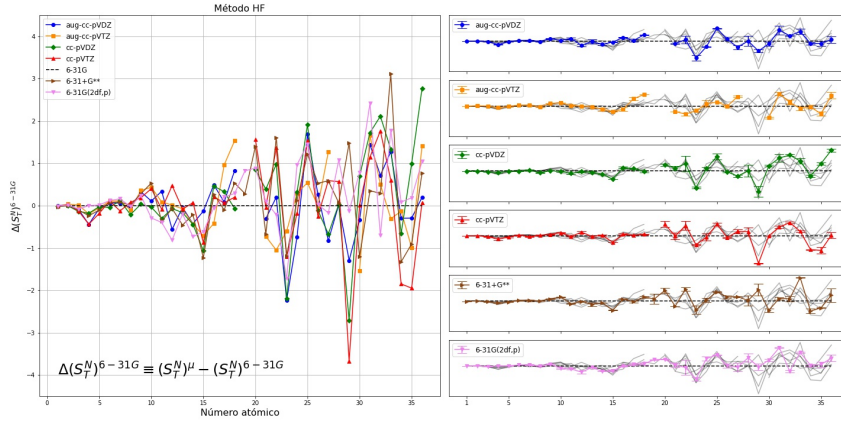
## Capítulo 4

# Resultados

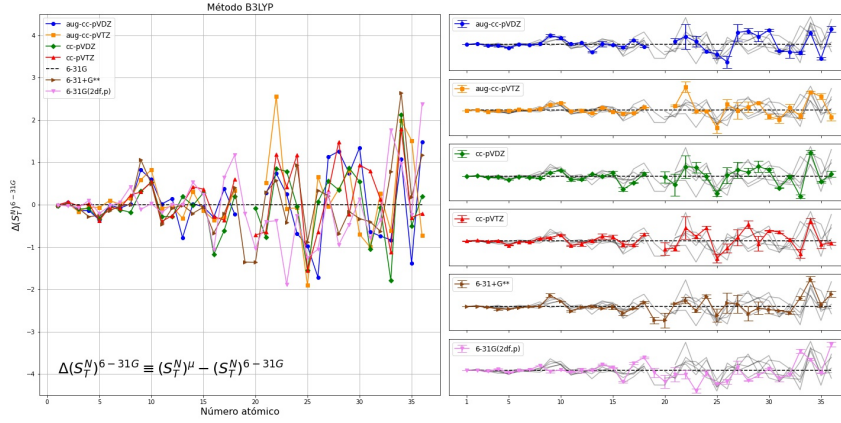
Las figuras 4.1 y 4.2 muestran las cantidades  $\Delta(S_T^{N_e})^{6-31G}$  para los métodos Hartree-Fock, B3LYP y  $\omega$ B97X-D, así como MP2(full), MP3(full) y MP4(SDQ), respectivamente, en todas las bases, tomando como referencia 6-31G, para átomos del hidrógeno al criptón y las barras de error presentes del lado derecho de las mismas figuras muestran precisiones de órdenes  $10^{-5}$  a  $10^{-1}$ , con una longitud de  $3\sigma$ , donde  $\sigma$  es la desviación estándar.

El comportamiento de las  $\Delta(S_T^{N_e})^{6-31G}$  para todas las bases que no son referencia muestran aparentes fluctuaciones alrededor del cero que van incrementando proporcionalmente con el número atómico, lo cual indica que las bases habrán de tener calidades semejantes para números atómicos cercanos a 1, mientras que discrepan notoriamente para aquellos cercanos a 36. Particularmente, estas fluctuaciones incrementan bruscamente a partir del número atómico 19, donde comienza la presencia de los metales de transición. Todo esto es debido a que los efectos relativistas influyen más en la descripción del comportamiento electrónico de los átomos conforme incrementa el número atómico y los métodos o funcionales utilizados no involucran tales efectos. A partir de los metales de transición, los efectos relativistas son mucho más notorios. Por otro lado, todas estas aparentes fluctuaciones son diferentes para cada método y funcional, lo cual indica que hay una dependencia mutua entre base y método (o funcional).

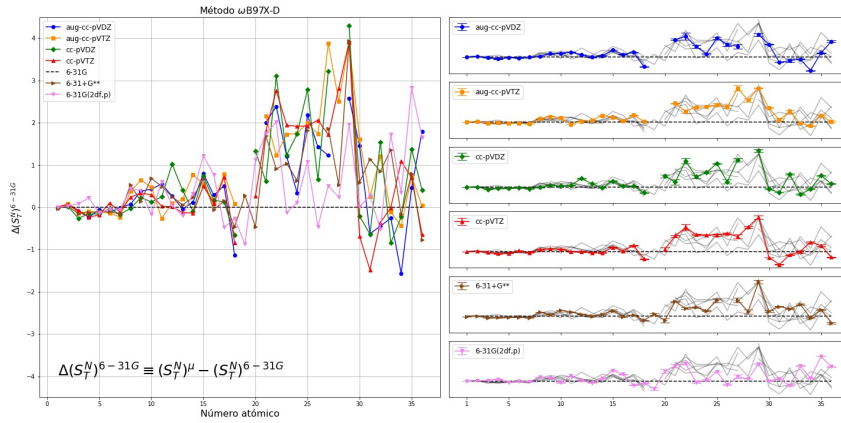
Por último, Los resultados mostrados en las figuras 4.1 y 4.2 muestran que el uso de la suma de la entropía de Shannon  $S_T^{N_e}$  podría no fungir como medida confiable para la calidad de las bases si se compara entre diferentes átomos, por ejemplo, el átomo con número atómico 21 en la figura 4.2a mostraría que la calidad en la base cc-pVTZ es mayor a aquella en la base 6-31G, mientras que el átomo con número atómico 22 indicaría el caso contrario (tabla 4.1). Particularmente, las bases con desdoblamiento triple  $\zeta$ , las cuales mejor describen una función de onda, tampoco presentarían una mejora constante contra las de desdoblamiento doble  $\zeta$ , por ejemplo, en la figura 4.1c, el átomo con número atómico 22 con las bases aug-cc-pVDZ y aug-cc-pVTZ (tabla 4.2a) y el átomo 27 de la figura 4.2b con las bases cc-pVDZ y cc-pVTZ (tabla 4.2b). Los resultados son semejantes si se compara entre bases con funciones difusas y bases que no las consideran, como son los casos del átomo 29 de la figura 4.2c, con las bases cc-pVTZ y aug-cc-pVTZ (tabla 4.2c), y del átomo 34 de la figura 4.1b, con las bases cc-pVDZ y aug-cc-pVDZ (tabla 4.2d). Por último, los resultados tampoco muestran diferencias constantes entre las bases 6-31G, 6-31+G\*\* y 6-31G(2df,p), es decir, para algunos átomos, la base 6-31G sería la mejor, como es el caso del átomo 22 de la figura 4.2a, y para otros la peor, como en el átomo 24 (tabla 4.3a), también sucede de forma semejante para las bases 6-31+G\*\*, como los átomos 15 y 33 de la figura 4.1a (tabla 4.3b), y 6-31G(2df,p) con los átomos 31 y 32 de la misma figura (tabla 4.3c).



(a)



(b)



(c)

Figura 4.1: Diferencias de entropía  $\Delta(S_T^{N_e})^{6-31G} \equiv (S_T^{N_e})^\mu - (S_T^{N_e})^{6-31G}$ , para las bases  $\mu \in \{\text{aug-cc-pVDZ, aug-cc-pVTZ, cc-pVDZ, cc-pVTZ, 6-31G, 6-31+G^{**}, 6-31G(2df,p)}\}$  contra número atómico (1–36), para los métodos HF (4.1a), B3LYP (4.1b) y  $\omega$ B97X-D (4.1c). El lado izquierdo muestra una comparación entre todas las bases mencionadas con referencia en 6-31G, mientras que el derecho se enfoca en cada una de ellas, agregando barras de error con longitud de  $3\sigma$  para cada uno de los valores obtenidos, donde  $\sigma$  es la desviación estándar. Se utilizaron 200,000 puntos de muestreo por cada base y átomo.

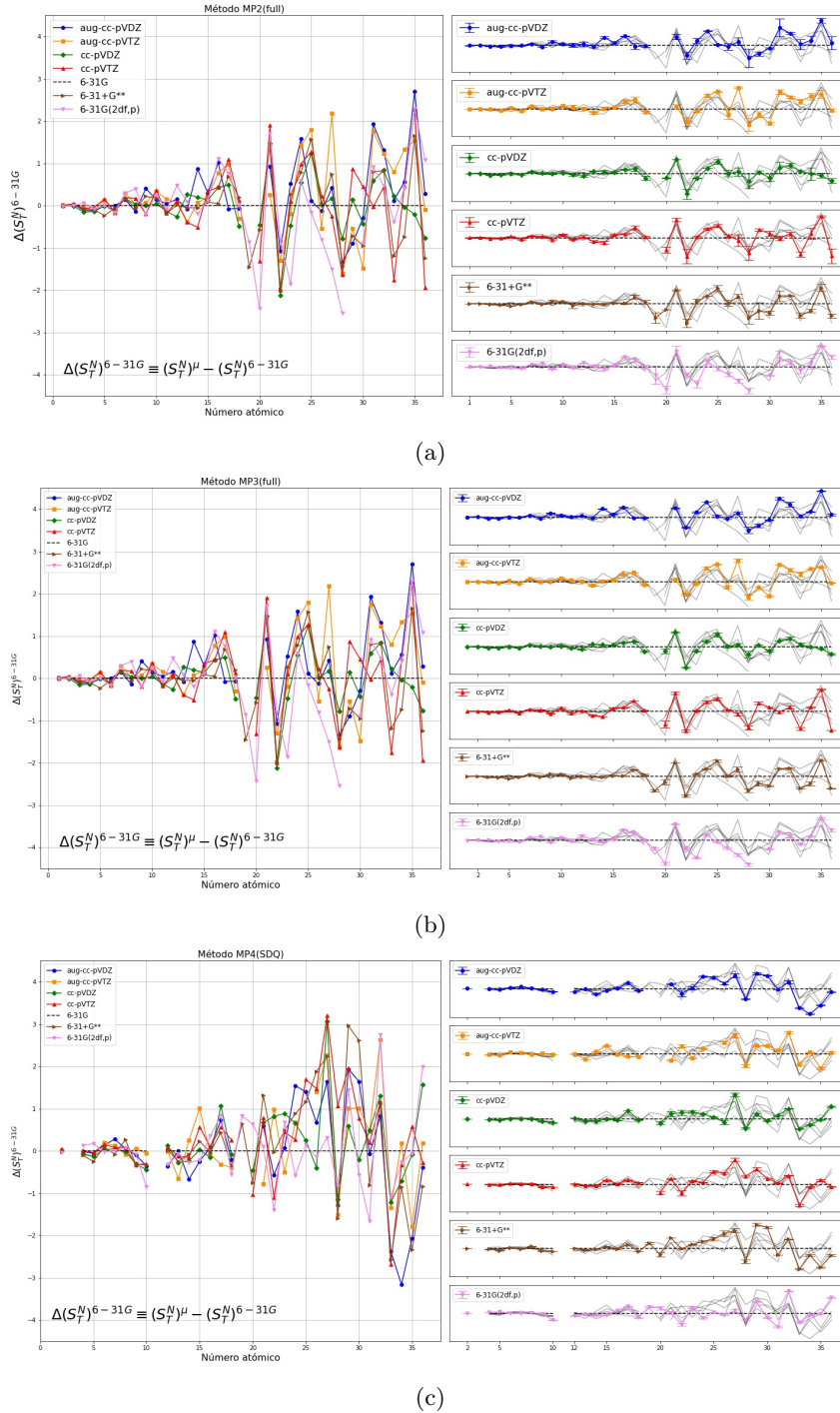


Figura 4.2: Diferencias de entropía  $\Delta(S_T^{N_e})^{6-31G} \equiv (S_T^{N_e})^\mu - (S_T^{N_e})^{6-31G}$ , para las bases  $\mu \in \{\text{aug-cc-pVDZ, aug-cc-pVTZ, cc-pVDZ, cc-pVTZ, 6-31G, 6-31+G^{**}, 6-31G(2df,p)}\}$  contra número atómico (1–36), para los métodos MP2(full) (4.2a), MP3(full) (4.2b) y MP4(SDQ) (4.2c). El lado izquierdo muestra una comparación entre todas las bases mencionadas con referencia en 6-31G, mientras que el derecho se enfoca en cada una de ellas, agregando barras de error con longitud de  $3\sigma$  para cada uno de los valores obtenidos, donde  $\sigma$  es la desviación estándar. Se utilizaron 200,000 puntos de muestreo por cada base y átomo.

Z	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
21	1.9065	0.0634
22	-1.9909	0.2599

Tabla 4.1: Comparación entre  $\Delta(S_T^{N_e})^{6-31G}$  para los átomos con números atómicos 21 y 22, ordenados de forma descendente. El método y la base utilizados fueron MP2(full) y cc-pVTZ.

Base	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
aug-cc-pVDZ	2.3872	0.1118
aug-cc-pVTZ	1.2313	0.1406

(a) Método  $\omega$ B97X-D, átomo 22

Base	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
cc-pVTZ	1.9628	0.0711
aug-cc-pVTZ	1.0075	0.131

Base	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
cc-pVDZ	3.3272	0.1552
cc-pVTZ	1.522	0.1709

(b) Método MP3(full), átomo 27

Base	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
cc-pVDZ	2.1194	0.0666
aug-cc-pVDZ	1.0728	0.0692

(c) Método MP4(SDQ), átomo 29

(d) Método B3LYP, átomo 34

Tabla 4.2: Comparación entre bases con desdoblamiento doble zeta y triple zeta para los diferentes átomos 22 (4.2a), 27 (4.2b), 29 (4.2c) y 34 (4.2d), así como métodos  $\omega$ B97X-D, MP3(full), MP4(SDQ) y B3LYP, respectivamente.

Z	Base	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
22	6-31G	0	0
	6-31G(2df,p)	-0.807	0.0653
24	cc-pVDZ	0.5555	0.0023
	6-31G	0	0

(a) Método MP2(full)

Z	Base	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
33	6-31+G**	3.1113	0.0071
	6-31G(2df,p)	1.775	0.0339
15	cc-pVDZ	-1.0601	0.0007
	6-31+G**	-1.2307	0.0126

(b) Método HF

(a) Método MP2(full)

(b) Método HF

Z	Base	$\Delta(S_T^{N_e})^{6-31G}$	$3\sigma$
31	6-31G(2df,p)	2.4186	0.0243
	cc-pVDZ	1.7215	0.1585
32	6-31+G**	0.2895	0.1082
	6-31G(2df,p)	-0.7009	0.0462

(c) Método HF

Tabla 4.3: Comparación de las bases 6-31G (4.3a), 6-31+G\*\* (4.3b) y 6-31G(2df,p) (4.3c) contra sus respectivas bases más cercanas, donde los métodos usados fueron MP2(full) y HF.

## Capítulo 5

# Conclusiones

Se desarrolló e implementó el algoritmo Las Vegas+ para funciones 3D sobre el paquete de programas *DensToolKit*, permitiendo utilizar tal algoritmo en átomos para analizar sus propiedades. Una vez implementado en *DensToolKit*, se analizó su eficiencia comparando sus cálculos con aquellos realizados por el método Monte Carlo, así como con aquellos publicados por Flores-Gallegos, N. [15] y Hô, M. et al. [2]. Se probó que el método implementado presenta una eficiencia muy por encima de Monte Carlo, sus cálculos mostraron buena precisión con no más de 200,000 puntos de muestreo y fueron comparables con los realizados por los mismos autores.

Se calcularon las entropías de Shannon sobre 36 átomos crecientes en número atómico y con métodos y bases distintos para corroborar una relación entre la entropía de Shannon y la calidad de las bases. Los resultados mostraron una aparente relación entre la entropía y los métodos y bases utilizados en el presente trabajo al notar fluctuaciones alrededor de la base 6-31 en las figuras 4.1 y 4.2 que van incrementando conforme crece el número atómico, sin embargo, no parece fungir adecuadamente como medida de la calidad de las bases, pues todas estas mostraban ser mejores para algunos átomos y peores para otros. Cabe mencionar que tales aparentes fluctuaciones se comportaron de forma distinta entre todas ellas, lo cual también mostró una dependencia entre método (o funcional) y base, además de que el método también influye en la entropía.

Se tienen esperados más trabajos para realizar análisis más detallados sobre la calidad de las bases a través de la entropía. Se propone comparar las entropías informacionales con las energías de los sistemas formados por distintas bases y calculadas por distintos métodos para determinar si hay una relación entre la entropía y la energía calculada numéricamente. También se pretende realizar un estudio semejante con las entropías informacionales propuestas por otros autores como Flores-Gallegos, N. [15] u otros y realizar el mismo estudio sobre moléculas para determinar si se cumplen comportamientos semejantes al de los átomos. Por último, dado que los métodos y funcionales influyen en la entropía informacional de átomos, se propone realizar un estudio donde se comparen diferentes métodos o funcionales por cada base.



# Apéndice A

## Validación del programa *dtkintegrate*

### A.1. Eficiencia de *dtkintegrate*

Para analizar la precisión del programa *dtkintegrate* sobre cualquier átomo, se calcularon los errores relativos de las integrales de las densidades electrónicas en los espacios de posiciones y momentos para cada uno de los átomos del hidrógeno al kriptón, y se promedió sobre ellos conforme se incrementaba el número de puntos de muestreo desde  $2 \times 10^5$  hasta  $2 \times 10^6$ . Los resultados obtenidos se muestran en la figura A.1 y en ellos se puede observar que los cálculos no superan el 1.1% de error relativo para las densidades electrónicas de los primeros 36 átomos de la tabla periódica.

Los parámetros utilizados para la integración fueron:

- Número de intervalos: 10.
- Máximo número de iteraciones: 20.
- Taza de convergencia: 1.
- Termalización: 0.
- No detener refinamiento de la malla durante la integración.
- Tolerancia: 0.

### A.2. Eficiencia de *dtkintegrate* contra el método Monte Carlo

También se puede probar la eficiencia del programa. Para eso, se desarrolló e implementó el algoritmo de integración Monte Carlo en *DensToolKit* y se compararon ambos programas, tomando como referencia el porcentaje de error relativo contra el número de evaluaciones sobre el integrando. La función a integrar fue la densidad electrónica del metano, cuyo número de electrones es 10 y el resto de los parámetros del programa *dtkintegrate* se mantuvieron fijos en los establecidos por defecto:

- Taza de convergencia: 1.
- Número de intervalos: 10.
- Máximo número de iteraciones: 20.

APÉNDICE A. VALIDACIÓN DEL PROGRAMA *DTKINTEGRATE*  
A.2. EFICIENCIA DE *DTKINTEGRATE* CONTRA EL MÉTODO MONTE CARLO

---

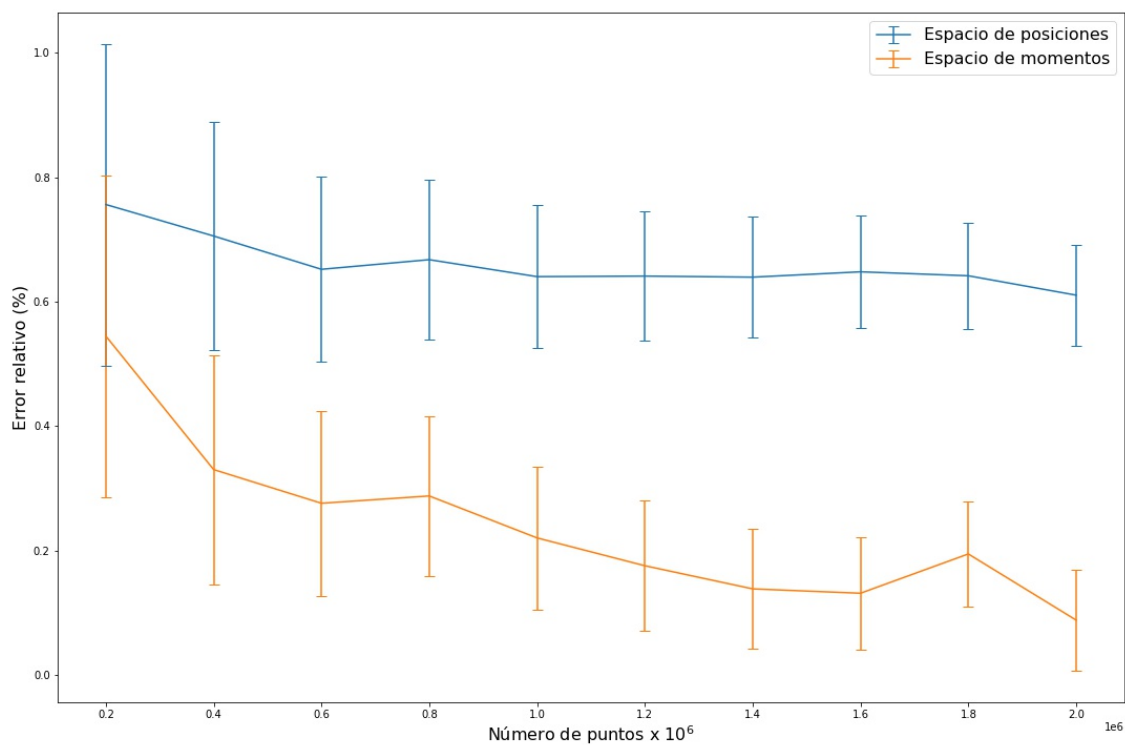


Figura A.1: Error relativo del promedio de las integrales de las densidades electrónicas de los átomos del hidrógeno al kriptón, tanto en los espacios de posiciones (azul) y momentos (naranja), contra número de puntos de muestreo ( $2 \times 10^5$  a  $2 \times 10^6$ ).

## APÉNDICE A. VALIDACIÓN DEL PROGRAMA *DTKINTEGRATE*

### A.3. ENTROPÍAS DE SHANNON EN ÁTOMOS UTILIZANDO *DTKINTEGRATE*

- Termalización: 0.
- No detener refinamiento de la malla durante la integración.
- Tolerancia: 0.

Se puede apreciar en la figura A.2 que los resultados obtenidos por Las Vegas+ son notablemente más acertados que aquellos por Monte Carlo, sin importar el número de evaluaciones que se realicen (puntos de muestreo) sobre el integrando.

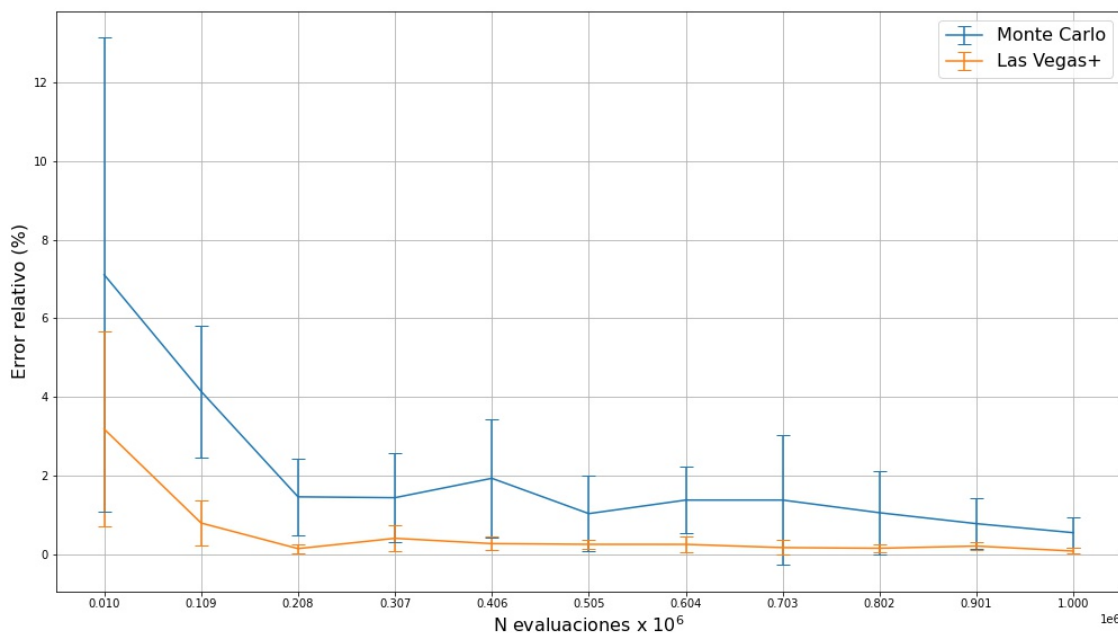


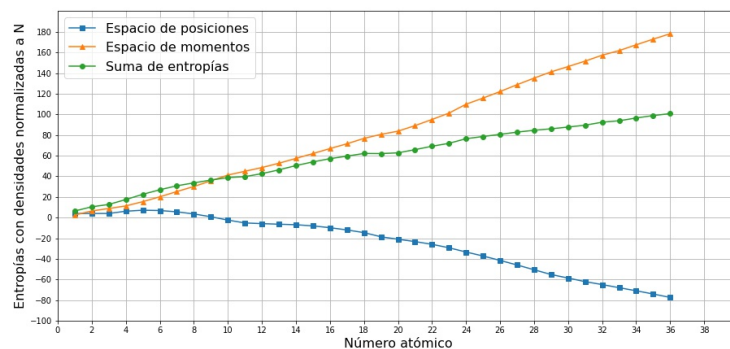
Figura A.2: Comparación del porcentaje de error relativo contra número de evaluaciones sobre la integral de la densidad electrónica del metano (10 electrones). Los métodos de integración usados fueron Monte Carlo (azul) y Las Vegas+ (naranja). Se usaron entre 10,000 y 1,000,000 evaluaciones sobre el integrando.

### A.3. Entropías de Shannon en átomos utilizando *dtkintegrate*

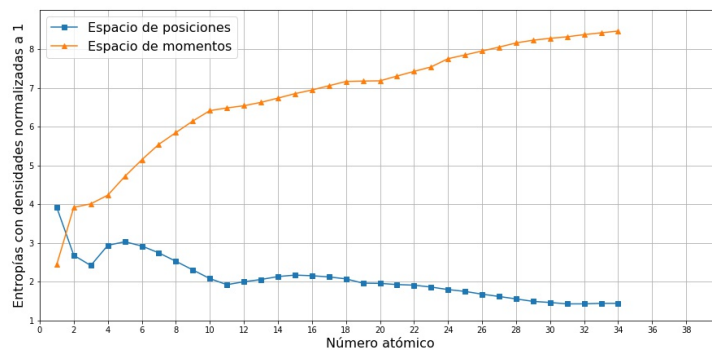
Para corroborar el funcionamiento del programa implementado en *DensToolKit*, se calcularon las entropías informacionales de las densidades normalizadas al número de electrones y a la unidad, así como de la densidad relativa, la cual toma como referencia su máximo global. La entropía se integró en átomos simples con números atómicos desde 1 hasta 36, utilizando como combinaciones método-base MP2/6-31+G\* y HF/cc-pvTZ, ambos obtenidos por GAUSSIAN09 [1]. Los resultados se compararon con aquellos publicados por Flores-Gallegos, N. [15] y Hô, M. et al. [2], quienes realizaron los mismos cálculos y con los mismos átomos, pero con combinaciones método-base CISD(full)/cc-pvTZ y HF/6-31G\*\*, respectivamente. Los resultados mostrados en las figuras A.3 presentan las mismas tendencias a aquellos publicados por los autores mencionados anteriormente.

Los resultados mencionados anteriormente muestran que el programa *dtkintegrate* se implementó adecuadamente en la suite *DensToolKit* y muestran una eficiencia adecuada de parte del algoritmo Las Vegas+ para funciones de comportamiento gaussiano y con centros cuasi-singulares.

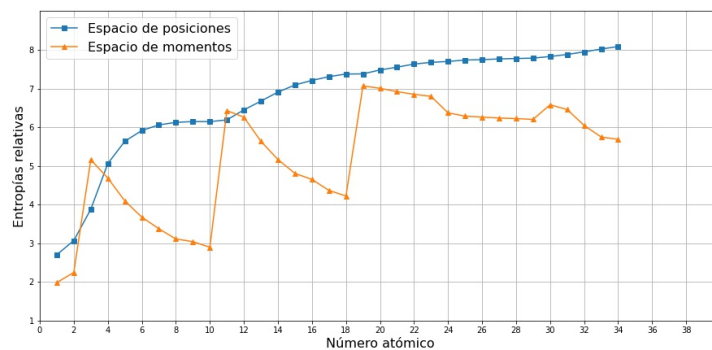
APÉNDICE A. VALIDACIÓN DEL PROGRAMA *DTKINTEGRATE*  
A.3. ENTROPÍAS DE SHANNON EN ÁTOMOS UTILIZANDO *DTKINTEGRATE*



(a) Entropías de Shannon sobre átomos desde el hidrógeno hasta el kriptón. La densidad electrónica está normalizada al número de electrones del átomo respectivo ( $N$ ) y se muestra en los espacios de posiciones (cuadrado azul) y momentos (triángulo naranja). Además, se presenta la suma de ambas entropías (círculo verde).



(b) Entropías de Shannon sobre átomos desde el hidrógeno hasta el selenio. La densidad electrónica está normalizada a la unidad y se muestra en los espacios de posiciones (cuadrado azul) y momentos (triángulo naranja).



(c) Entropías de Shannon sobre átomos desde el hidrógeno hasta el selenio. La densidad electrónica es relativa y toma como referencia el valor máximo de la densidad de cada átomo. También se muestra en los espacios de posiciones (cuadrado azul) y momentos (triángulo naranja).

Figura A.3

# Bibliografía

- [1] K. Remya and C. H. Suresh, “Which density functional is close to ccSD accuracy to describe geometry and interaction energy of small noncovalent dimers? a benchmark study using gaussian09,” *Journal of computational chemistry*, vol. 34, no. 15, pp. 1341–1353, 2013.
- [2] M. Hô, R. P. Sagar, J. M. Pérez-Jordá, V. H. Smith Jr, and R. O. Esquivel, “A numerical study of molecular information entropies,” *Chemical Physics Letters*, vol. 219, no. 1-2, pp. 15–20, 1994.
- [3] M. Hô, R. P. Sagar, V. H. Smith Jr, and R. O. Esquivel, “Atomic information entropies beyond the hartree-fock limit,” *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 27, no. 21, p. 5149, 1994.
- [4] A. N. Tripathi, R. P. Sagar, R. O. Esquivel, and V. H. Smith Jr, “Electron correlation in momentum space: The beryllium-atom isoelectronic sequence,” *Physical Review A*, vol. 45, no. 7, p. 4385, 1992.
- [5] S. R. Gadre, S. B. Sears, S. J. Chakravorty, and R. D. Bendale, “Some novel characteristics of atomic information entropies,” *Physical Review A*, vol. 32, no. 5, p. 2602, 1985.
- [6] G. Peter Lepage, “A new algorithm for adaptive multidimensional integration,” *Journal of Computational Physics*, vol. 27, no. 2, pp. 192–203, 1978.
- [7] G. Grasseau, S. Lisniak, and D. Chamont, “Hybrid implementation of the vegas monte-carlo algorithm,” *Proceedings, GPU Computing in High-Energy Physics (GPUHEP2014): Pisa, Italy*, pp. 103–108, 2014.
- [8] G. P. Lepage, “Vegas: An adaptive multi-dimensional integration program,” *Cornell Preprint CLNS*, pp. 80–447, 1980.
- [9] S. Carrazza and J. M. Cruz-Martinez, “Vegasflow: accelerating monte carlo simulation across multiple hardware platforms,” *Computer Physics Communications*, vol. 254, p. 107376, 2020.
- [10] T. Ohl, “Vegas revisited: Adaptive monte carlo integration beyond factorization,” *Computer Physics Communications*, vol. 120, no. 1, pp. 13–19, 1999.
- [11] G. P. Lepage, “Adaptive multidimensional integration: vegas enhanced,” *Journal of Computational Physics*, vol. 439, p. 110386, 2021.
- [12] C. E. Shannon, “A mathematical theory of communication,” *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, p. 3–55, Jan. 2001.
- [13] T. M. Cover, J. A. Thomas, *et al.*, “Entropy, relative entropy and mutual information,” *Elements of Information Theory*, vol. 2, no. 1, pp. 12–13, 1991.
- [14] F. Jensen, *Introduction to computational chemistry*. John Wiley & Sons, 2017.

- [15] N. Flores-Gallegos, "On the calculations of shannon's entropy in atoms and molecules i: The continuous case in position and momentum spaces," *Chemical Physics Letters*, vol. 720, pp. 1–6, 2019.
- [16] I. Białyński-Birula and J. Mycielski, "Uncertainty relations for information entropy in wave mechanics," *Communications in Mathematical Physics*, vol. 44, no. 2, pp. 129–132, 1975.
- [17] E. Clementi and C. Roetti, "Roothaan-hartree-fock atomic wavefunctions: Basis functions and their coefficients for ground and certain excited states of neutral and ionized atoms, *zle54*," *Atomic data and nuclear data tables*, vol. 14, no. 3-4, pp. 177–478, 1974.
- [18] G. BERTHIER, "Extension de la methode du champ moleculaire self-consistent a letude des etats a couches incompletes," *Comptes Rendus Hebdomadaires des Seances De l'Academie des Sciences*, vol. 238, no. 1, pp. 91–93, 1954.
- [19] R. McWeeny and G. Dierksen, "Self-consistent perturbation theory. ii. extension to open shells," *The Journal of Chemical Physics*, vol. 49, no. 11, pp. 4852–4856, 1968.
- [20] J. A. Pople and R. K. Nesbet, "Self-consistent orbitals for radicals," *The Journal of Chemical Physics*, vol. 22, no. 3, pp. 571–572, 1954.
- [21] C. C. J. Roothaan, "New developments in molecular orbital theory," *Reviews of Modern Physics*, vol. 23, no. 2, p. 69, 1951.
- [22] D. A. McQuarrie, *Quantum chemistry*. University Science Books, 2008.
- [23] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus, "New basis set exchange: An open, up-to-date resource for the molecular sciences community," *Journal of Chemical Information and Modeling*, vol. 59, no. 11, pp. 4814–4820, 2019.
- [24] M. J. Frisch, M. Head-Gordon, and J. A. Pople, "A direct mp2 gradient method," *Chemical Physics Letters*, vol. 166, no. 3, pp. 275–280, 1990.
- [25] J. A. Pople, J. S. Binkley, and R. Seeger, "Theoretical models incorporating electron correlation," *International Journal of Quantum Chemistry*, vol. 10, no. S10, pp. 1–19, 1976.
- [26] R. Krishnan and J. A. Pople, "Approximate fourth-order perturbation theory of the electron correlation energy," *International Journal of Quantum Chemistry*, vol. 14, no. 1, pp. 91–100, 1978.
- [27] P. R. Surján, *The Brillouin Theorem*, pp. 87–92. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989.
- [28] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review*, vol. 140, no. 4A, p. A1133, 1965.
- [29] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *The Journal of Physical Chemistry*, vol. 98, no. 45, pp. 11623–11627, 1994.
- [30] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A*, vol. 38, no. 6, p. 3098, 1988.
- [31] C. Lee, W. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, no. 2, p. 785, 1988.
- [32] J.-D. Chai and M. Head-Gordon, "Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections," *Physical Chemistry Chemical Physics*, vol. 10, no. 44, pp. 6615–6620, 2008.

- [33] J.-D. Chai and M. Head-Gordon, "Systematic optimization of long-range corrected hybrid density functionals," *The Journal of Chemical Physics*, vol. 128, no. 8, p. 084106, 2008.
- [34] J. C. Slater, "Atomic shielding constants," *Physical Review*, vol. 36, no. 1, p. 57, 1930.
- [35] S. F. Boys, "Electronic wave functions-i. a general method of calculation for the stationary states of any molecular system," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 200, no. 1063, pp. 542–554, 1950.
- [36] S. Huzinaga, "Gaussian-type functions for polyatomic systems. i," *The Journal of Chemical Physics*, vol. 42, no. 4, pp. 1293–1302, 1965.
- [37] J. Solano-Altamirano and J. M. Hernández-Pérez, "Denstoolkit: A comprehensive open-source package for analyzing the electron density and its derivative scalar and vector fields," *Computer Physics Communications*, vol. 196, pp. 362–371, 2015.
- [38] M. Matsumoto and T. Nishimura, "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 8, no. 1, pp. 3–30, 1998.