[Thesis Title]

Thesis by

[Your Full Name]

In Partial Fulfillment of the Requirements for the

Degree of

[Name of Degree]

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

[Year Degree Conferred]

Defended [Exact Date]

Choose one from the choices in the source code!! And delete this `todo` when you're done that. :-)

ACKNOWLEDGEMENTS

[Add acknowledgements here. If you do not wish to add any to your thesis, you may simply add a blank titled Acknowledgements page.]

ABSTRACT

[This abstract must provide a succinct and informative condensation of your work. Candidates are welcome to prepare a lengthier abstract for inclusion in the dissertation, and provide a shorter one in the CaltechTHESIS record.]

PUBLISHED CONTENT AND CONTRIBUTIONS

[Include a bibliography of published articles or other material that are included as part of the thesis. Describe your role with the each article and its contents. Citations must include DOIs or publisher URLs if available electronically.

If you are incorporating any third-party material in the thesis, including works that you have authored/co-authored but for which you have transferred copyright, you must indicate that permission has been secured to use the material. For example: "Fig. 2 reprinted with permission from the copyright holder, holder name"

Add the option `iknowwhattodo` to this environment to dismiss this message.]

1. Cahn, J. K. B., Baumschlager, A., Brinkmann-Chen, S. & Arnold, F. H. Mutations in adenine-binding pockets enhance catalytic properties of NAD (P) H-dependent enzymes. *Protein Engineering Design and Selection* **19,** 31–38 (2016).
   J.K.B.C participated in the conception of the project, solved and analyzed the crystal structures, prepared the data, and participated in the writing of the manuscript.

2. Cahn, J. K. B., Brinkmann-Chen, S., Spatzal, T., Wiig, J. A., Buller, A. R., Einsle, O., Hu, Y., Ribbe, M. W. & H., A. F. Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases. *Biochemical Journal* **468,** 475–484 (2015).
   J.K.B.C participated in the conception of the project, solved and analyzed the crystal structures, prepared the data, and participated in the writing of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number   *Page*

LIST OF TABLES

Number *Page*

Chapter 1

NOMENCLATURE

This uses the restricted Hartree-Fock formalism, meaning that two electrons with opposite spin occupy each spatial orbital. All calculations have been done using the PySCF package.[1]

Symbol	Description
i, j, k, l	Occupied orbital indices
a, b, c, d	Virtual orbital indices
p, q, r, s	General MO indices
$\mu, \nu, \lambda, \sigma$	AO indices
$(pq\|rs) = \int \int \psi_p^*(\mathbf{r}_1)\psi_q(\mathbf{r}_1)\frac{1}{r_{12}}\psi_r^*(\mathbf{r}_2)\psi_s(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2$	Two-electron integrals
$(pq\|\|rs) = (pq\|rs) - (ps\|rq)$	Antisymmetrized two-electron integrals

C h a p t e r   2

MOTIVATION

The formalism of many-body perturbation theory (MBPT) can provide corrections to a mean-field description such as that given by Hartree-Fock or density functional theory (DFT). The former method does not treat electron correlation at all, and the latter one treats it in an average way. DFT is often used for systems of large size, as it is fairly accurate and computationally cheap. However, DFT's reliance on empirically based functionals can give rise to the notorious self-interaction error; because one is considering an electron in the average field of the electrons in the system, you can have an electron interacting with itself. In practice, this can lead to a variety of issues, including the underestimation of surface stability (overestimation of surface energies) relevant in surface science studies.[2] To remedy this problem, normally one would fall back onto the wave function-based MBPT methods, such as coupled cluster theory (CC) and Moller-Plesset perturbation theory to 2nd order (MP2). However, their computational scaling is steep for larger systems, such as periodic ones.[3] Because of this issue, the perturbative methods often are not a better option than the field-standard DFT in predicting the properties of materials, as I have learned in a previous study.[4] Therefore, in the quantum chemistry community, there has been an interest in applying Green's function MBPT methods, with the *GW* approximation, to such systems, which has shown to give accurate corrections to, for example, band gaps, on top of a prior (DFT) mean-field calculation, at a cheap computational cost.[5]

I have included a plot that shows the difference in the HOMO energy for the water molecule with different basis choices for the Green's function method, G_0W_0, and regular DFT, both with same starting point from the PBE functional. As can be seen, G_0W_0 captures more of the electronic correlation, thus lowering the orbital energy.

HOMO Energy vs. Basis Set for Water @PBE
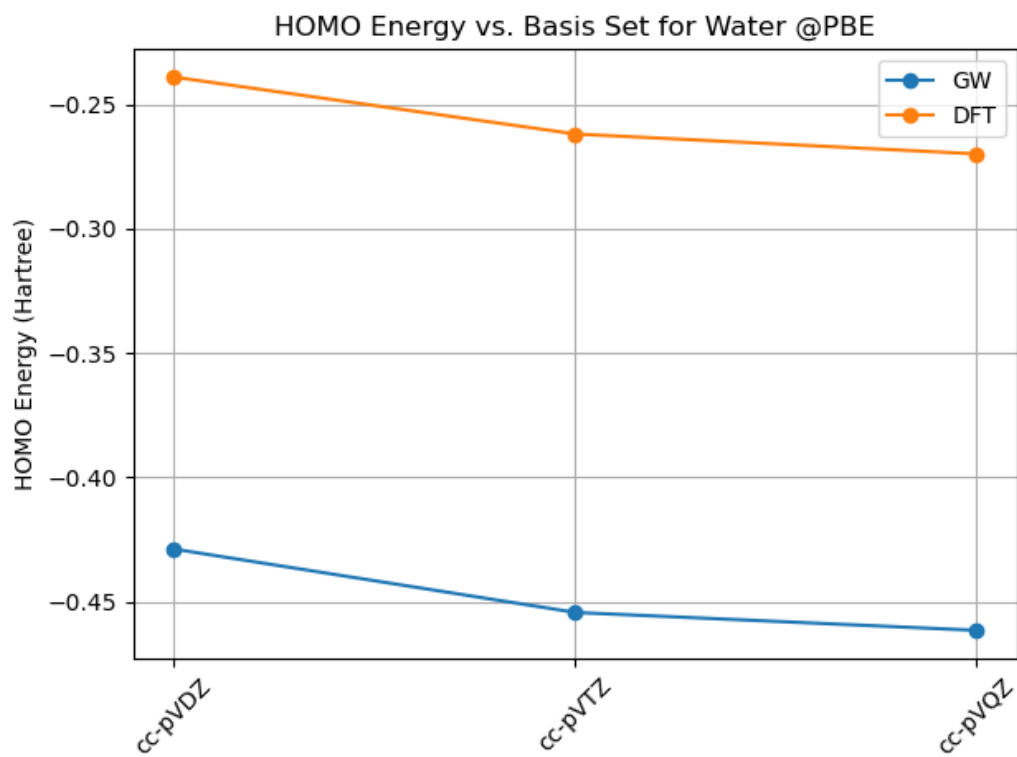
C h a p t e r   3

THEORETICAL BACKGROUND

We begin by writing out the Schrödinger equation for the N-electron system in a molecular field.

$$\hat{H}\Psi = E\Psi \tag{3.1}$$

with

$$\hat{H} = \sum_{i=1}^{N}\left(-\frac{1}{2}\nabla_i^2\right) - \sum_{i=1}^{N}\sum_{\alpha}\frac{Z_\alpha}{r_{i\alpha}} + \sum_{i<j}\frac{1}{r_{ij}} \tag{3.2}$$

We note that we do not include the potential energy term associated with the nuclear-nuclear repulsion, as we are working in the Born-Oppenheimer approximation. This can be succinctly written as the sum of a kinetic energy term with a nuclear-electronic attraction term and an electron-electron repulsion term

$$\hat{H} = \hat{T} + \hat{V}_{ne} + \hat{V}_{ee} \tag{3.3}$$

The objective of many years of research in quantum chemistry has been solving for the V_{ee} term. Now, I will introduce some of the mean field methods that have been classically used to tackle this problem at a cheap computational cost.

3.1   Mean Field Methods

Hartree-Fock

In the restricted Hartree-Fock formalism, the V_{ee} is given by

$$V_{ee} = \frac{1}{2}\sum_{i}^{occ}\sum_{j}^{occ} 2(ii|jj) - (ij|ji) \tag{3.4}$$

The first term is the Coulomb term and the second term is an exchange term. We choose to make the following presentation in the restricted Hartree-Fock formalism, but considering spin here is important. In the unrestricted Hartree-Fock formalism, this would be given by

$$V_{ee} = \frac{1}{2}\sum_{\underline{i}}^{occ}\sum_{\underline{j}}^{occ} \left((\underline{ii}|\underline{jj}) - (\underline{ij}|\underline{ji})\right) \tag{3.5}$$

Where the underline denotes that these are spin orbitals; we have not performed the summation over all possible spins yet. The Hartree-Fock method fails to take

into account the exchange term in the unrestricted case; the integral vanishes when the spins of i and j are different. Physically, this means that there is no energetic stabilization for two electrons with different spins to occupy the same spatial index, whereas there is one when the spins are parallel. This was what was meant earlier by the statement that Hartree-Fock does not treat electron correlation.

Density Functional Theory (DFT)

The central quantity in DFT is the electron density, as represented by a density matrix. The method is fairly black box, and its intricate details will not be covered here, but it is important to understand that it treats the electron correlation through an exchange-correlation functional, which in principle has an exact form, but in practice is approximated as semi-empirical, which leads to issues in accurately describing the electron correlation.

What is important to understand is that both of these mean field methods yield molecular orbital energies ϵ and coefficients $C_{\mu p}$, which we can correct with MBPT methods.

3.2 Many-body perturbation theory

There are many different MBPT methods, but they all have the same purpose of recovering a portion of the electron correlation missed by these mean field methods. For example, Hartree-Fock is variational, meaning that its energy E_{HF} will always be greater or equal to the true energy of the system E_0. For HF, we define the difference between these two quantities as the correlation energy

$$E_{\text{corr}} \equiv E_0 - E_{HF} \tag{3.6}$$

The other mean field method that is often used is DFT, but it is not variational, so there is not such a simple interpretation of E_{corr} in this case, but the principle is similar. We will be focusing on recovering E_{corr} with the G_0W_0 method, which is a Green's function-based MBPT method.

Chapter 4

G_0W_0 PROCEDURE

4.1 Iterative equation

The procedure that was used to compute quasiparticle energies, which can be interpreted as effective molecular orbital energies, is given by the below equation:

$$\delta_{pq}F_{pq}^{HF}[\gamma^{DFT}] + \Sigma^{corr}(\varepsilon_p^{QP}) = \varepsilon_p^{QP} \tag{4.1}$$

We explain the notation starting from left to right. The first term corresponds to taking the diagonal of the Hartree-Fock matrix F_{pq}^{HF} evaluated at a given electron density γ. These electron densities are obtained from a previous mean-field calculation, either γ_{DFT} or γ_{HF}. The one-electron reduced density matrix was used to express the electron density here. The second term evaluates Σ^{corr} for the ε_p^{QP} determined in the previous iteration. The right side of the equality gives the updated ε_p^{QP}.

4.1 is iterated until self-consistency. We start with an initial guess for ε_p^{QP}, which is given by the mean-field orbital energy ϵ_p. This is used in the first iteration to solve for the right-hand side ε_p^{QP} of 4.1. In the next iteration, we use the previously obtained ε_p^{QP} to formulate Σ^{corr}. This process is repeated until we reach a convergence threshold for ε_p^{QP}.

The Fock Matrix

In the basis of atomic orbitals, this is given by:

$$F_{\mu\nu}^{HF} = h_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma}(\mu\nu|\lambda\sigma) - \frac{1}{2}\sum_{\lambda\sigma} P_{\lambda\sigma}(\mu\lambda|\nu\sigma) \tag{4.2}$$

where $h_{\mu\nu}$ is the one-electron part of the Hamiltonian, $P_{\lambda\sigma}$ is the one-electron reduced density matrix, and $(\mu\nu|\lambda\sigma)$ is a two-electron integral.[6] A useful identity for $P_{\lambda\sigma}$ in terms of the MO coefficients from the mean-field calculation is:

$$P_{\mu\nu} = 2\sum_{i=1}^{N/2} C_{\mu i}C_{\nu i}^* \tag{4.3}$$

We note that the sum runs only over the $N/2$ occupied *spatial* orbitals. 4.2 is the form for the HF Fock matrix, and not the DFT Fock matrix. In the canonical HF MO

basis, $F_{\mu\nu}^{HF}$ merely has the HF MO energies placed on the diagonal. We transform the Fock matrix into the MO basis with:

$$F_{pq} = \sum_{\mu} \sum_{\nu} C_{\mu p}^{*} F_{\mu\nu} C_{\nu q} \tag{4.4}$$

where C is the matrix of MO coefficients.

Correlation-Self Energy

Σ^{corr} is the second term in 4.1. It is the central quantity in the GW method, and is formed from the inputted Green's function G and the screened Coulomb potential W. It is tasked with recovering a portion of the E_{corr} mentioned earlier. It is dynamic, as opposed to the previous Fock term that was discussed, as it is updated with a new ε_p^{QP} in each iteration. In the case of the G_0W_0 approximation, we use the common approximation, considering only the diagonal element of Σ^{corr} corresponding to the orbital with index p. This function is evaluated at the ε_p^{QP} just obtained in the previous iteration. To summarize, we are really interested in $\Sigma_{pp}^{corr}(\varepsilon_p^{QP})$. We will go into greater detail about the form of Σ^{corr} in the next chapter.

updated ε_p^{QP}

This is the right side, or the solution, of 4.1.

Graphical solution

One is able to check 4.1 in the HF case by plotting Σ^{corr} as a function of the input frequency, which takes on the possible values for ε_p^{QP}. Since F_{pq}^{HF} is diagonal in the canonical HF MO basis, 4.1 can be reformulated as

$$\epsilon_p^{HF} + \Sigma_{pp}^{corr}(\omega) = \omega \rightarrow \Sigma_{pp}^{corr}(\omega) = \omega - \epsilon_p^{HF} \tag{4.5}$$

Essentially, the line at $\omega - \epsilon_p^{HF}$ should, and does, intersect with Σ^{corr} at the same ε_p^{QP} that we get from our iterative procedure. This is a useful check to see if the self-energy, which we will derive in the next chapter, is being computed correctly.
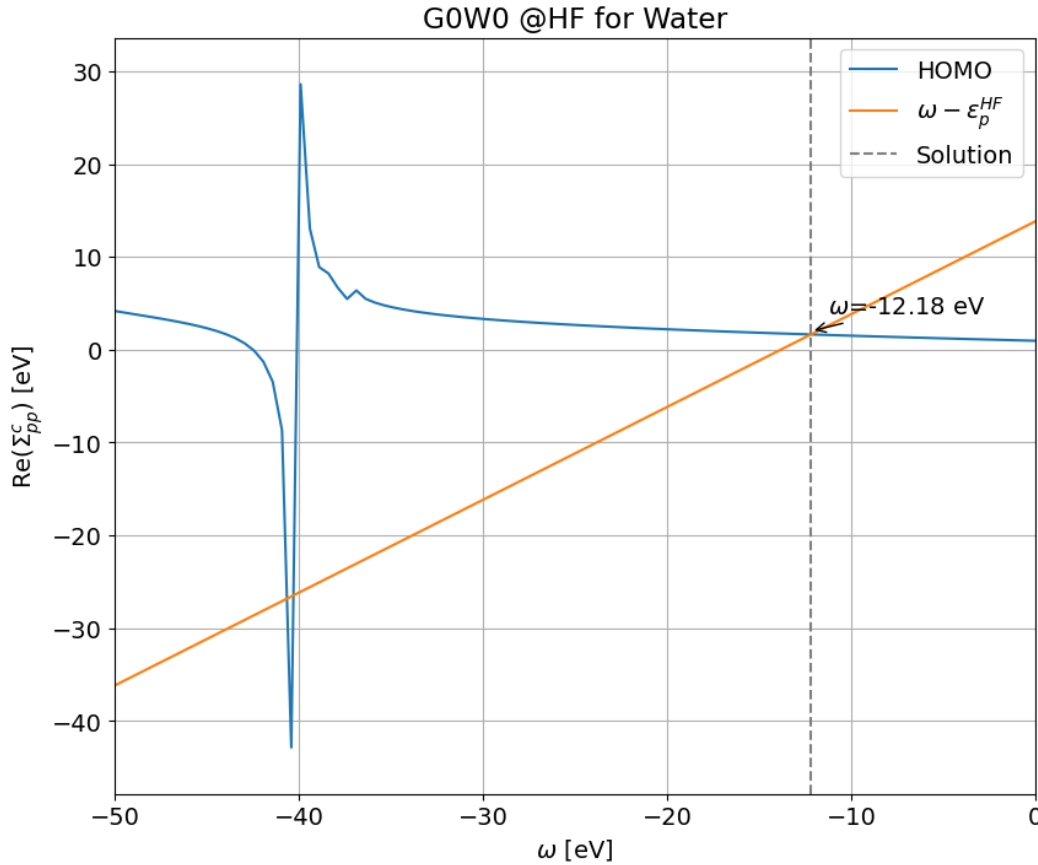
Also, at around ω = -40 eV, one can observe a pole structure. This would pose problems for my iterative procedure 4.1 if the ε_p^{QP} that I was looking for was close to this value.

4.2 Correlation Self-Energy

$$\Sigma_{pp}^{corr}(\omega) = \sum_{\mu}^{RPA} \left(\sum_{i}^{occupied} \frac{V_{pi}^{\mu} V_{ip}^{\mu}}{\omega - (\epsilon_i - \Omega_\mu)} + \sum_{a}^{virtual} \frac{V_{pa}^{\mu} V_{ap}^{\mu}}{\omega - (\epsilon_a + \Omega_\mu)} \right) \tag{4.6}$$

This is the working equation for the diagonal of the correlation self-energy for a given MO. The V^μ and Ω_μ are the excitation vectors and energies, respectively, from a previous RPA calculation. ω is my input frequency and the ϵ are the orbital energies from my previous mean-field calculation.

4.3 Time dependent DFT

Random Phase Approximation

The RPA is a linear response theory that is used to compute the excitation energies and vectors. The working matrix equation is given by[7]:

$$\begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \tag{4.7}$$

where A is

$$\mathbf{A}_{ia,jb} = \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) + 2(ia||jb) \tag{4.8}$$

and B is

$$\mathbf{B}_{ia,jb} = 2(ia||jb) \tag{4.9}$$

The excitation vectors $\mathbf{V}^\mu$ are taken by considering a contraction of two tensors. First, we consider the sum of $\mathbf{X}$ and $\mathbf{Y}$ at the same excitation energy μ: $\mathbf{Z}_{i,a,\mu} = \mathbf{X}_{i,a,\mu} + \mathbf{Y}_{i,a,\mu}$. Then we contract this with the two-electron integrals:

$$\mathbf{W}_{p,q,i,a} = \sqrt{2} \sum_{p,q,i,a} (pq|ia) \tag{4.10}$$

This factor of $\sqrt{2}$ comes from the spin integration of the restricted Hartree-Fock formalism. We defined a combined occupied-virtual index ν, so: $\mathbf{Z}_{i,a,\mu} \rightarrow \mathbf{Z}_{\nu,\mu}$ and $\mathbf{W}_{p,q,i,a} \rightarrow \mathbf{W}_{p,q,\nu}$.

And then we form the excitation vector from:

$$\mathbf{V}^\mu_{pq} = \sum_\nu \mathbf{W}_{p,q,\nu} \mathbf{Z}_{\nu,\mu} \tag{4.11}$$

Tamm-Dancoff Approximation

In this method, we neglect the $\mathbf{B}$ matrix of the RPA equation. So the eigenvalue equation becomes

$$\mathbf{AX} = \omega\mathbf{X} \tag{4.12}$$

where we still have:

$$\mathbf{A}_{ia,jb} = \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) + 2(ia||jb) \tag{4.13}$$

And then we follow the same procedure as in the RPA to get $\mathbf{V}^{\mu}_{pq}$, where now we have $\mathbf{Z}_{\nu,\mu} = \mathbf{X}_{\nu,\mu}$.

Direct approximation

Everywhere in the code, we consider the direct approximation, which just means that all instances of anti-symmetrized two-electron integrals are replaced by their non-symmetrized counterparts. In 4.8, 4.9, and 4.13, $(ia||jb) \rightarrow (ia|jb)$. In the former case it was called the direct Random Phase Approximation (dRPA) and in the latter case it was called the direct Tamm-Dancoff Approximation (dTDA).

C h a p t e r   5

LINEARIZED G_0W_0 DENSITY MATRIX

Implementation

These are the working equations for the linearized G_0W_0 Density Matrix that I will derive later.[8] First, we consider the fully occupied block:

$$\gamma_{ij}^{GW} = 2\delta_{ij} - 2 \sum_{a\mu} \frac{\mathbf{V}_{ia}^{\mu} \mathbf{V}_{ja}^{\mu}}{\left(\epsilon_i - \epsilon_a - \Omega_\mu\right)\left(\epsilon_j - \epsilon_a - \Omega_\mu\right)} \tag{5.1}$$

where the Ω_μ are the excitation energies and the $\mathbf{V}^\mu$ are the excitation vectors. The sum runs over all virtual orbitals and all excitation energies. The ϵ are the orbital energies from the prior mean-field calculation. Next, we have the virtual-virtual block:

$$\gamma_{ab}^{GW} = -2 \sum_{i\mu} \frac{\mathbf{V}_{ai}^{\mu} \mathbf{V}_{bi}^{\mu}}{\left(\epsilon_i - \epsilon_a - \Omega_\mu\right)\left(\epsilon_i - \epsilon_b - \Omega_\mu\right)} \tag{5.2}$$

Finally, we have the mixed block:

$$\gamma_{ib}^{GW} = \frac{2}{\epsilon_i - \epsilon_b} \left[\sum_{a\mu} \frac{\mathbf{V}_{ia}^{\mu} \mathbf{V}_{ba}^{\mu}}{\epsilon_i - \epsilon_a - \Omega_\mu} - \sum_{j\mu} \frac{\mathbf{V}_{ij}^{\mu} \mathbf{V}_{bj}^{\mu}}{\epsilon_j - \epsilon_b - \Omega_\mu} \right] \tag{5.3}$$

This all contributes to the form of the density matrix as:

$$2 \begin{pmatrix} \gamma_{ij}^{GW} & \gamma_{ib}^{GW} \\ \gamma_{bi}^{GW} & \gamma_{ab}^{GW} \end{pmatrix} \tag{5.4}$$

Where γ_{bi}^{GW} is simply the transpose of $\gamma_{ib}^{\mathrm{GW}}$, since all elements of this matrix are real. Therefore, this density matrix is Hermitian. The factor of 2 comes from the fact that we sum over both spins in the restricted Hartree-Fock formalism.

Plotting natural occupations

The natural occupations can be found by the diagonalizing the density matrix. They can be interpreted as being the amount of electrons in a given orbital.[6] Here we considered the one-electron density matrix from multiple methods; We started with Hartree-Fock, which contains no correlation, then we considered our implementation of the direct Random Phase Approximation (dRPA) and the direct Tamm-Dancoff Approximation (dTDA). Finally, we considered Full Configuration Interaction (FCI) as a reference, as it contains the exact correlation. It should be
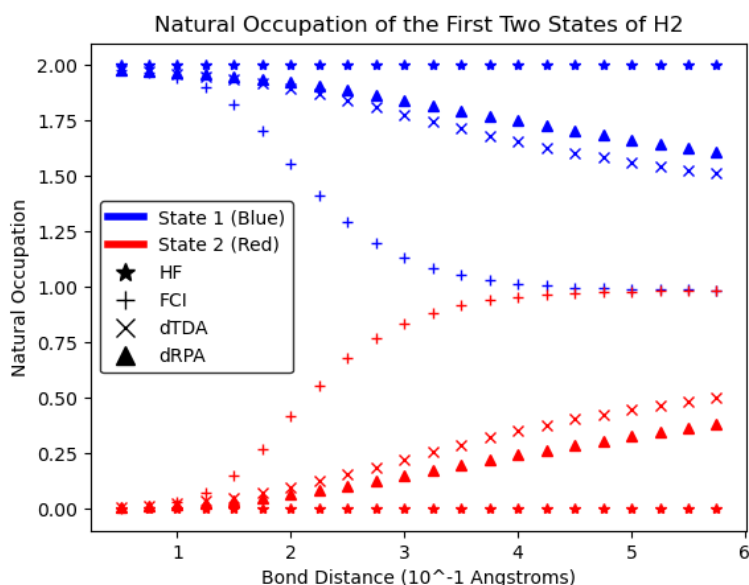
noted that we considered natural occupations of the HOMO (State 1) and LUMO (State 2) of H_2, which has the simple MO diagram.
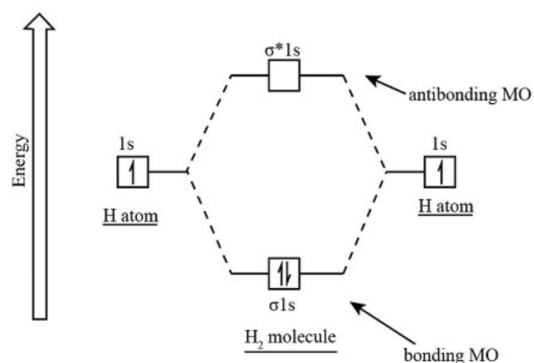
Chapter 6

THIS IS THE SIXTH CHAPTER

Chapter 7

THIS IS THE SEVENTH CHAPTER

Chapter 8

THIS IS THE EIGHTH CHAPTER

BIBLIOGRAPHY

1. Sun, Q. *et al.* Recent developments in the PySCF program package. *The Journal of Chemical Physics* **153,** 024109. ISSN: 0021-9606. `https://doi.org/10.1063/5.0006074` (2024) (July 2020).

2. Schimka, L. *et al.* Accurate surface and adsorption energies from many-body perturbation theory. en. *Nature Materials* **9.** Publisher: Nature Publishing Group, 741–744. ISSN: 1476-4660. `https://www.nature.com/articles/nmat2806` (2024) (Sept. 2010).

3. McClain, J., Sun, Q., Chan, G. K.-L. & Berkelbach, T. C. Gaussian-Based Coupled-Cluster Theory for the Ground-State and Band Structure of Solids. *Journal of Chemical Theory and Computation* **13.** Publisher: American Chemical Society, 1209–1218. ISSN: 1549-9618. `https://doi.org/10.1021/acs.jctc.7b00049` (2024) (Mar. 2017).

4. Kozlowski, P. Elucidating Catalysis with the "Gold Standard" of Quantum Chemistry. en. *Caltech Undergraduate Research Journal.* `https://curj.caltech.edu/2021/06/29/elucidating-catalysis-with-the-gold-standard-of-quantum-chemistry/` (2024) (June 2021).

5. *Frontiers | The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy* `https://www.frontiersin.org/articles/10.3389/fchem.2019.00377/full` (2024).

6. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* en. Google-Books-ID: KQ3DAgAAQBAJ. ISBN: 978-0-486-13459-8 (Courier Corporation, June 2012).

7. Dreuw, A. & Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chemical Reviews* **105.** Publisher: American Chemical Society, 4009–4037. ISSN: 0009-2665. `https://doi.org/10.1021/cr0505627` (2024) (Nov. 2005).

8. Bruneval, F. Assessment of the Linearized GW Density Matrix for Molecules. *Journal of Chemical Theory and Computation* **15.** Publisher: American Chemical Society, 4069–4078. ISSN: 1549-9618. `https://doi.org/10.1021/acs.jctc.9b00333` (2024) (July 2019).

Appendix A

QUESTIONNAIRE

Appendix B

CONSENT FORM

POCKET MATERIAL: MAP OF CASE STUDY SOLAR
SYSTEMS