

# Computational analysis of vibrational frequencies using CNDO/2 with STO-3G basis sets

*Beshay P., McTighe J.  
CHEM 279 Final Project Fall 2025*

## **Abstract**

Vibrational modes of molecules provide valuable insight into how said molecules move in space. These insights can help scientists to better understand kinetics and provide a means of identification through the analysis of frequencies at which the bonds expand and contract. In this project, a method was developed for analyzing these modes and frequencies using Complete Neglect of Differential Overlap (CNDO) with Slater Type Orbital (STO) basis sets. From an initial geometry file, the positions were iteratively optimized, and the vibrational frequencies calculated. Overall, the results diverged significantly from empirical values. This divergence is most likely based on poorly chosen basis sets and insufficient geometric optimization, which has been well studied in prior work. Investigating these faults provided a better understanding of the theory behind these analytical methods and how to design a better package going forward.

## Background

Molecular spectroscopy, particularly the study of vibrational modes via infrared (IR) spectroscopy, serves as a powerful analytical tool for characterizing chemical systems and understanding molecular dynamics. The unique set of vibrational frequencies for a given molecule acts as a “fingerprint”, providing characteristic information about its structure, bonding, and how atoms move about their equilibrium positions. These physical observations can be modeled in a computational setting using computational methods. The theoretical foundation for these analyses often relies on the calculation of the potential energy surface (PES) and subsequent normal-mode analysis (NMA), typically modeled as a harmonic oscillator. This approach requires the computation of the second energy derivative matrix (Hessian) at a stationary point on the PES.

Computational chemistry methods have emerged to complement and aid the interpretation of experimental data by providing theoretical expressions that link observable measurements to underlying physical properties. Early efforts in computational spectroscopy sought to balance computational efficiency with accuracy, leading to the development of semi-empirical methods. Integrals deployed in semi-empirical methods are either determined directly from experimental data or calculated from a corresponding analytical formula represented by a suitable parametric expression [1].

One such method, CNDO/2, was developed by Pople to generalize early theoretical approaches for valence electrons and ensure rotational invariance [1]. The original parameterization of the CNDO method, such as CNDO/2, was designed to reproduce ab initio Hartree-Fock results obtained with a minimal basis set, often the STO-3G (Slater Type Orbitals approximated by 3 Gaussian functions) basis set. The STO-3G is a minimal basis set, meaning it uses the fewest possible functions to describe the electrons in an atom. While computationally inexpensive, these early methods and minimal basis sets were known to have limitations. Their primary aim was often qualitative understanding rather than quantitative prediction of precise experimental values, and they often produced results that differed significantly from empirical data [2]. While methods such as Density Functional Theory (DFT) are generally preferred for their higher accuracy, methods like CNDO can provide rapid qualitative information for understanding chemical systems, as they are approximately 1000 times faster than DFT methods [1].

Mathematically, the Hessian matrix is derived from the PES with respect to position. The force constants are then weighted by the masses of the pair of elements represented in each entry of the matrix. Following a diagonalizing step, the vibrational frequencies and vibrational normal coordinates can be calculated from the eigenvalues and eigenvectors, respectively. In order for the force constants contained within the Hessian to be accurate, geometric position must be optimized. This reduces interference from other forms of molecular motion, such as rotation and translation. In simulations with un-optimized geometries, these different modes will “mix” in with the vibrational modes and produce inaccurate results [3]. Theoretically, a completely optimized geometry will result in a gradient value of zero on the PES. Because these optimizations are constrained by computational power, this gradient value will never be exactly zero, so some interference from the other modes will be expected. An additional constraint,

imposed by the Born-Oppenheimer approximation, leads to ambiguities in the vibrational mass, thus vibrational frequencies have an accuracy limit of  $\sim 0.1 \text{ cm}^{-1}$  [4].

## Objectives

The objectives of this project were multi-fold, focusing on the implementation and critical evaluation of a computational methodology for analyzing molecular vibrational modes. The primary objective was to create a functional system that takes a raw geometry file, optimizes its structure using steepest descent, and calculates the corresponding vibrational frequencies. The project aimed to utilize the CNDO approximation to understand the practical application and computational efficiency of established semi-empirical quantum chemistry methods in predicting molecular properties. A key objective in this implementation was to systematically compare the computationally derived vibrational frequencies and IR plots with established empirical values to quantify the accuracy and inherent limitations in the chosen CNDO/STO-3G methodology. Related to these limitations, this study sought to investigate and analyze potential sources of significant divergence between theoretical and empirical results, specifically addressing the impact of the chosen minimal basis sets and the thoroughness of the geometric optimization process.

## Methodology

Several different molecules and increasing levels of complexity were utilized to explore the accuracy and limitations of utilizing CNDO/2 to derive IR frequencies. These included H<sub>2</sub>, HF (Linear diatomic molecules), H<sub>2</sub>O (3 atom geometry), NH<sub>3</sub> (4 atom) and CH<sub>4</sub> (5 atom). The workflow was divided into two general steps, Geometry Optimization and calculation of the IR frequencies. Container-Native development was utilized throughout the project for consistent deployment and testing across systems. A docker container based on HW5 was utilized as a skeleton equipped with computational dependencies such as OpenBLAS, LAPACK, ARPACK, Armadillo & Eigen.

### Geometry Optimization

As mentioned previously, it is critical to use molecules that are at their optimal geometry/lowest energy state because a less stable geometry would cause noise that can translate into inaccurate force constants & Hessians with negative eigenvalues (imaginary frequencies) indicating unstable modes.

In order to optimize the geometry of each molecule, a steepest-descent algorithm was implemented to identify the optimal geometry. The procedure can be summarized in the following steps:

1. Parse the molecule file to get atom position and the number of alpha & beta electrons.
2. Build Contracted Gaussian basis sets (STO-3G) from these atoms.
3. Compute initial total energy of the molecule at the starting geometry.
  - a. Using CNDO/2 Self-Consistent Field(SCF), specifically Unrestricted Hartree-Fock (UHF).

$$E_{\text{CNDO/2}} = \frac{1}{2} \sum_{\mu\nu} p_{\mu\nu}^{\alpha} \left( h_{\mu\nu} + f_{\mu\nu}^{\alpha} \right) + \frac{1}{2} \sum_{\mu\nu} p_{\mu\nu}^{\beta} \left( h_{\mu\nu} + f_{\mu\nu}^{\beta} \right) + \sum_A \sum_{B < A} \frac{Z_A Z_B}{R_{AB}}$$

4. Compute the analytic CNDO/2 gradient & gradient norm (check convergence state).

$$E_{\text{CNDO/2}}^{\mathbf{R}_A} = \sum_{\mu \neq \nu} x_{\mu\nu} s_{\mu\nu}^{\mathbf{R}_A} + \sum_{B \neq A} y_{AB} \gamma_{AB}^{\mathbf{R}_A} + V_{\text{nuc}}^{\mathbf{R}_A}$$

5. Perform a Line Search by moving against force gradient using bracketing & golden section minimization to determine optimal step size.
6. Steps 3-5 are repeated until the state converges, meaning the difference between gradients in steps is below the specified tolerance (default 1.0 e -6).
  - a. Energy calculations after the first iteration re-use previous density/Fock matrices as initial guesses as opposed to a zero filled matrix.
  - b. The loop also breaks once set max iterations are reached or 5 consecutive iterations result in no change.

### Vibrational Analysis

Once optimal geometry is achieved, the next step is to construct the Hessian-matrix by calculating the second derivatives of the calculated total energy (using SCF as the previous step) with respect to positions. This matrix is then mass-weighted and diagonalized in order to obtain the vibrational frequencies & normal modes.

1. Parse molecule file to get atom position and number of alpha & beta electrons.
2. Build Contracted Gaussian basis set (STO-3G)
3. Construct Hessian Matrix via central finite differences.
  - a. Default step size of 0.001.
4. Calculate analytic gradient.
  - a. Provides the force vector for each atom.
  - b. Gradient is used to construct the Hessian (Force Constant  $\mathbf{F}$ )

$$V(\Delta\mathbf{x}) = \frac{1}{2} \Delta\mathbf{x}^t \mathbf{F} \Delta\mathbf{x}$$

5. Mass-Weight Hessian.
  - a. Convert Cartesian coordinates to mass-weighted coordinates ( $\mathbf{G}$  Matrix).

$$y_i = \sqrt{m_i} \Delta x_i \quad ; \quad \frac{\partial^2}{\partial y_i^2} = \frac{1}{m_i} \frac{\partial^2}{\partial x_i^2} \quad ; \quad G_{ij} = \frac{1}{\sqrt{m_i m_j}}$$

6. Diagonalize Mass-Weighted Hessian (Solve eigenvalue problem), outputs:
  - a. Eigenvalues(Frequencies)
  - b. Eigenvectors (vibrational normal coordinates)

## Results

**Table 1.** Comparison of total energy (electronic + repulsion) per molecule before and after steepest descent optimization with 100 iterations.

Molecule	Initial Energy (eV)	SD Optimized Energy (eV)	Delta
H <sub>2</sub>	-26.75	-30.78	-4.03
HF	-171.5505	-179.4359	-7.8854
H <sub>2</sub> O	-156.0088	-169.2565	-13.2477
NH <sub>3</sub>	-141.3439	-155.4033	-14.0594
CH <sub>4</sub>	-137.9968	-157.3192	-19.3224

**Table 2.** Bond stretch frequency outputs for molecules with unoptimized geometries, optimized with steepest descent (SD) and pre-optimized geometries from the Materials Project (MP)

Frequencies and Percentage Difference From Empirical values (Bond Stretch Modes)					
Molecule	H <sub>2</sub>	HF	H <sub>2</sub> O	NH <sub>3</sub>	CH <sub>4</sub>
(Unoptimized) Calculated cm <sup>-1</sup>	4042	6117	Not Measured	Not Measured	Not Measured
(SD) Calculated cm <sup>-1</sup>	8140	1162	4228, 4393	Not Measured	Not measured
(MP) Calculated cm <sup>-1</sup>	5911	7242	6580, 6596	4587, 4785	4619, 4730
Calculated (MP Scaled) cm <sup>-1</sup>	3994	4893	4446, 4457	3099, 3233	3121, 3196
Empirical cm <sup>-1</sup>	4342 [5]	4460 [6]	3506, 3585 [7]	3464, 3534 [8]	3104, 3215 [9]
% Difference (MP vs Empirical)	30.6%	47.55%	60.98%, 59.15%	27.9%, 30.07%	39.2%, 38.15%
% Difference (Scaled MP vs Empirical)	8.0%	9.26%	23.63%, 21.71%	11.12%, 8.90%	0.55%, 0.59%

While the steepest descent optimization results were initially promising, with all molecules reaching gradient below threshold within a few iterations, utilizing these optimized geometries yielded vibrational modes that deviated significantly from the empirical values as shown in **Table 1**. For the linear molecules, the optimized geometries yielded worse outputs than the unoptimized geometries. Overall, the highest accuracy was achieved when utilizing molecules with pre-optimized geometries sourced from the Materials Project. The scaling factor recommended by Shaw et al [2] typically used for outputs derived from this computation method (1.48) yielded better results for all molecules (H<sub>2</sub> 30.6% vs 8.0%, HF 47.55% vs 9.26% etc.). There is also a potential trend showing that the most symmetrical molecule CH<sub>4</sub> had the highest accuracy across all results with the scaling factor (0.55%, 0.59% difference). More investigation is required to see if the model performing better for molecules with large degrees of symmetry in their structure is a trend or if other variables caused CH<sub>4</sub> to have higher accuracy.

## Discussion

The geometry optimization phase, utilizing the steepest descent algorithm, was largely unsuccessful due to significant computational expense and convergence issues. The resulting geometries, as visualized for H<sub>2</sub> and H<sub>2</sub>O, showed notable discrepancies when compared to established geometries from the Materials Project. This failure is primarily attributed to the limitations of the simple STO-3G basis set, which often underestimates bond lengths, and the model's inability to fully account for electron correlations by only considering valence electrons. Modern semi-empirical methods, such as MP2 or DFT are generally considered more effective for geometry calculations.

The calculated vibrational frequencies also exhibited poor accuracy when compared to empirical values, a deficiency that was partially mitigated by applying a scaling factor of 1.48 as recommended by Shaw et al. Review of existing literature pointed to two main sources of error: the basis set and un-optimized geometries. The STO-3G basis set was identified by Shaw et al as "completely unreliable" for modeling vibrational frequencies, as it significantly overestimates frequencies due to its inflexibility [2]. Alternative basis sets like 3-21G or 4-31G are suggested substitutes [2,3]. The 3-21G is a split valence basis set, where the core orbitals are represented by a contraction of 3 Gaussian functions, the inner valence orbitals are represented by a contraction of 2 Gaussians, and the outer valence is represented by 1 Gaussian. While this contains the same number of primitive Gaussians as an STO-3G basis set, it is much more flexible because there are twice as many valence functions that can combine freely to represent a molecular orbital [4].

For the purpose of vibrational calculations, an ideal basis set would be one that accounts for the changing electron density as bonds in a molecule expand and contract. This property is described as being "polarization consistent". Such basis sets are often used in DFT calculations as they describe the polarization of electron density upon the formation of a molecule, rather than the correlation energy between atoms [4]. Examples of polarization consistent functions being DZP and TZP type basis sets.

Additionally, the presence of notably high rotational frequencies (up to  $\sim$ 500 cm<sup>-1</sup>) in the eigenvalues clearly indicates sub-optimal molecular geometries, leading to a mixing of rotational and true vibrational results that propagates inaccuracies. In addition to utilizing geometry optimization to ensure the gradient is at a stationary point, transforming the geometries to normalized coordinates may also be employed to reduce rotational and translation interference. However, this projection can only occur at stationary

geometries. At non-stationary geometries, a set of generalized frequencies may be obtained by removing the gradient direction from the force constants via projection techniques [4]. Implementing such normalization techniques in this work would likely reduce interference from molecular rotation and improved results.

## Conclusion

This study sought to perform geometric optimization of molecules and calculate their associated vibrational frequencies using CNDO/2 coupled with STO-3G basis sets. Despite the low numerical accuracy of the results, the project provided a valuable opportunity to expand knowledge on the theoretical and practical limitations of computational chemistry methodologies. The discrepancies necessitated a deeper investigation into the underlying CNDO/2 approximations, the constraints of the chosen STO-3G basis set, and the practical challenges of achieving optimal geometries. This process emphasized the critical role of selecting appropriate mathematical models and basis sets to effectively represent a target application in computational modeling.

## Contributions

Philo Beshay - Developed Dockerfile for container-native development of the system based on the HW5 skeleton. Created steepest geometry optimization functionality as a precursor to vibration analysis, including Pybind module for python interfacing and visualization. Gathered & formatted data sets to be used for testing (from HW5 & Materials Project).

James McTighe - Developed CMake build system for C++ source files. Organized headers from previous homework assignments into a modular format for re-use. Designed core vibrational analysis system. Created bash scripts for building source files and running them using files contained in the MolJSON directory.

## Future Work

As suggested in the discussion section, the developed package lacked sufficient accuracy because of the chosen basis set and poor geometry optimization. The initial implementation of the optimizer was far more computationally taxing, significant efficiency improvements were achieved by passing the previous SCF state with each iteration to avoid repeating the full calculation at each iteration. Further efficiency improvements may be possible by refactoring code to store additional state in between iterations. Most importantly, in order for this package to be useful, new functionality for different basis sets must be implemented for the calculations. Based on study by Shaw et al, 3-21G would be an appropriate next step. For enhanced code modularity, this could be done as a plugin where the end user supplies their own design for a basis set, allowing for maximum re-usability of existing code. Additionally, bindings can also be implemented for the IR spec executable for more seamless python interfacing especially using directly in Jupyter notebooks. One of the most immediate next steps would be to analyze additional larger molecules (6+ atoms) in order to investigate if the accuracy increasing with larger molecules is a true trend or if other variables were responsible for the accuracy difference.

## References

- (1) Thiel, W. Semiempirical Quantum–Chemical Methods. *WIREs Computational Molecular Science* 2013, 4 (2), 145–157. <https://doi.org/10.1002/wcms.1161>.
- (2) Shaw, R. A.; Ursenbach, C.; Rauk, A.; Wieser, H. Comparison of STO-3G and 3-21G Ab Initio Harmonic Force Fields for Ethane, Propane, Dimethyl Ether, and Cyclobutane: Effects of Geometry and Scaling on Calculated Frequencies, Eigenvectors, and Infrared Absorption Intensities. *Canadian Journal of Chemistry* 1988, 66 (5), 1318–1332. <https://doi.org/10.1139/v88-214>.
- (3) Bernardi, F.; Schlegel, H. B.; Wolfe, S. Ab Initio Computation of Force Constants. *Journal of Molecular Structure* 1976, 35 (1), 149–153. [https://doi.org/10.1016/0022-2860\(76\)80110-x](https://doi.org/10.1016/0022-2860(76)80110-x).
- (4) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons, 2016.
- (5) Nash, J. J. *Vibrational Modes of Hydrogen ( $H_2$ )*; Purdue University JMol Vibrational Data. Purdue University Department of Chemistry. <https://www.chem.purdue.edu/jmol/vibs/h2.html> (accessed December 2025). Data based on HyperChem AM1 calculations. [Purdue Chemistry+1](#)
- (6) Nash, J. J. *Vibrational Modes of Hydrogen Fluoride (HF)*; Purdue University JMol Vibrational Data. Purdue University Department of Chemistry. <https://www.chem.purdue.edu/jmol/vibs/h2.html> (accessed December 2025). Data based on HyperChem AM1 calculations. [Purdue Chemistry](#)
- (7) Nash, J. J. *Vibrational Modes of Water ( $H_2O$ )*; Purdue University JMol Vibrational Data. Purdue University Department of Chemistry. <https://www.chem.purdue.edu/jmol/vibs/h2o.html> (accessed December 2025). Data based on HyperChem AM1 calculations. [Purdue Chemistry+1](#)
- (8) Nash, J. J. *Vibrational Modes of Ammonia ( $NH_3$ )*; Purdue University JMol Vibrational Data. Purdue University Department of Chemistry. <https://www.chem.purdue.edu/jmol/vibs/nh3.html> (accessed December 2025). Data based on HyperChem AM1 calculations. [Purdue Chemistry](#)
- (9) Nash, J. J. *Vibrational Modes of Methane ( $CH_4$ )*; Purdue University JMol Vibrational Data. Purdue University Department of Chemistry. <https://www.chem.purdue.edu/jmol/vibs/ch4.html> (accessed December 2025). Data based on HyperChem AM1 calculations.

## Appendices

Github repository: [James-McTighe/chem\\_279\\_final\\_project](https://github.com/James-McTighe/chem_279_final_project)

**Figure 1.** Geometry Comparison for Steepest Descent vs Pre-Optimized Molecules (Blue:Unoptimized, Green:SD, Red: Pre-Optimized from Materials Project)

Molecule	Unoptimized Geometry	Optimized (CNDO/2 Steepest Descent)	Optimized (Materials Project)
H <sub>2</sub>			
HF			
HO			
H <sub>2</sub> O			
CH <sub>3</sub>			

**Figure 2.A.** MP\_Optimized H2 Full IR Output

Hessian					
9.68E-03	0	0	-9.68E-03	0	0
8.47E-21	2.29E-04	0	-8.47E-21	-2.29E-04	0
7.22E-21	0	2.29E-04	-7.22E-21	0	-2.29E-04
-9.68E-03	0	0	9.68E-03	0	0
8.47E-21	-2.29E-04	0	-8.47E-21	2.29E-04	0
7.22E-21	0	-2.29E-04	-7.22E-21	0	2.29E-04

  

Frequency					
0					
0					
0					
613.85					
613.85					
3994.1					

**Figure 2.B.** MP\_Optimized HF Full IR Output

Hessian						
2.76E-02	-5.80E-15	0	-6.35E-03	4.45E-16	8.91E-16	
1.64E-15	-1.45E-03	-2.21E-16	-2.42E-16	3.33E-04	2.45E-17	
1.78E-20	-6.71E-33	-1.45E-03	-4.11E-21	7.73E-34	3.33E-04	
-6.35E-03	1.34E-15	-4.45E-16	1.46E-03	-1.03E-16	-2.05E-16	
-3.78E-16	3.33E-04	5.09E-17	5.56E-17	-7.67E-05	-5.64E-18	
-2.37E-20	-6.18E-33	3.33E-04	5.47E-21	0	-7.67E-05	

  

Frequency						
0						
0						
0						
0						
0						
4.89E+03						

**Figure 2.C.** MP\_Optimized H<sub>2</sub>O Full IR Output

Hessian						
1.49E-03	-8.48E-05	-2.98E-04	-5.55E-03	-1.11E-04	-3.86E-04	
-8.48E-05	-2.33E-05	3.71E-04	-5.49E-05	2.39E-04	-5.76E-05	
-2.98E-04	3.71E-04	1.17E-03	-1.90E-04	-5.76E-05	5.40E-05	
-5.55E-03	-5.49E-05	-1.90E-04	2.26E-02	1.66E-04	5.71E-04	
-1.11E-04	2.39E-04	-5.76E-05	1.66E-04	-9.75E-04	2.54E-04	
-3.86E-04	-5.76E-05	5.40E-05	5.71E-04	2.54E-04	-1.58E-04	
Frequency						
0						
0						
0						
0						
0						
1.09E+00						
8.63E+02						
4.45E+03						
4.46E+03						

**Figure 2.D.** MP\_Optimized NH3 Full IR Output

Hessian														
1.37E-03	5.51E-08	8.51E-08	-1.08E-03	-1.12E-03	-1.04E-07	-1.03E-03	1.10E-03	-9.95E-08	-2.99E-03	2.63E-05	-1.14E-07			
5.51E-08	1.37E-03	3.82E-10	-1.12E-03	-2.32E-03	-1.24E-07	1.10E-03	-2.36E-03	1.22E-07	2.63E-05	-4.14E-04	7.14E-10			
8.51E-08	3.82E-10	-9.03E-05	-9.43E-08	-1.08E-07	1.11E-04	-9.02E-08	1.05E-07	1.11E-04	-1.33E-07	9.05E-10	1.14E-04			
-1.08E-03	-1.12E-03	-9.43E-08	4.08E-03	4.39E-03	3.24E-07	1.50E-04	2.43E-04	4.23E-08	-2.06E-04	-4.48E-04	-1.48E-08			
-1.12E-03	-2.32E-03	-1.08E-07	4.39E-03	8.92E-03	4.33E-07	-2.52E-04	-3.18E-04	-3.50E-08	4.66E-05	3.67E-05	3.90E-09			
-1.04E-07	-1.24E-07	1.11E-04	3.24E-07	4.33E-07	3.09E-04	4.25E-08	3.42E-08	-3.62E-04	2.06E-08	-5.19E-09	-3.62E-04			
-1.03E-03	1.10E-03	-9.02E-08	1.50E-04	-2.52E-04	4.26E-08	3.90E-03	-4.29E-03	3.08E-07	-1.98E-04	4.53E-04	-1.41E-08			
1.10E-03	-2.36E-03	1.05E-07	2.43E-04	-3.18E-04	3.42E-08	-4.29E-03	9.09E-03	-4.24E-07	-4.20E-05	2.85E-05	-2.95E-09			
-9.95E-08	1.22E-07	1.11E-04	4.22E-08	-3.50E-08	-3.62E-04	3.08E-07	-4.24E-07	3.10E-04	2.06E-08	5.42E-09	-3.62E-04			
-2.99E-03	2.63E-05	-1.33E-07	-2.06E-04	4.66E-05	2.05E-08	-1.98E-04	-4.20E-05	2.05E-08	1.16E-02	-1.03E-04	4.53E-07			
2.63E-05	-4.14E-04	9.05E-10	-4.48E-04	3.67E-05	-5.10E-09	4.53E-04	2.85E-05	5.34E-09	-1.03E-04	1.48E-03	-3.61E-09			
-1.14E-07	7.14E-10	1.14E-04	-1.48E-08	3.79E-09	-3.62E-04	-1.41E-08	-2.84E-09	-3.62E-04	4.53E-07	-3.61E-09	2.99E-04			
Frequencies														
0														
0														
8.20E-01														
1.05E+00														
7.40E+02														
7.42E+02														
7.44E+02														
1.30E+03														
1.31E+03														
3.10E+03														
3.23E+03														
3.23E+03														

**Figure 2.E.** MP\_Optimized CH4 Full IR Output