Thermochemistry of Hydrocarbons: A Computational Approach Using the MTB/2 Method

Dulce Torres & Sabrina Temesghen

1 Abstract

To estimate the heat of formation and total energies of hydrocarbons, we implement Voityuk's MTB/2, a modified Extended Hückel approach. An improved accuracy compared to the extended huckel method was expected due to the consideration of electron-electron repulsion. However, comparison to experimental values reported a lack of accuracy in the applied MTB/2. The best results were obtained for aromatic structures. On the other hand, larger deviations from experimental values were observed in branched alkanes and strained structures.

2 Background

The heat of formation, or enthalpy, is a key thermodynamic parameter which serves as a quantitative description for chemical reactions. While experimental enthalpy values can be obtained from chemistry databases, values are often missing for larger sized systems. Alternatives to missing experimental values are the application of quantum chemical methods. However, these methods are computationally demanding and lose accuracy with increasing system size.[7] In order to avoid these limitations, the MTB/2 method was the applied method for determining the relative heat of formation and total energy of extended hydrocarbons.

3 Objectives

The purpose of this project is to implement and evaluate the MTB/2 method for determining the heat of formation and total energy for several different hydrocarbons including alkanes, unsaturated aliphatic, cyclic, and aromatic compounds. Total energy and enthalpy values were calculated with respect to experimental benzene and H_2, respectively.

4 Methodology

4.1 H matrix

MTB/2 assumes an orthogonalized atomic orbital (AO) basis, this implies there is no overlap between basis functions of different atoms, thereby simplifying the matrix construction process. The Hamiltonian Matrix, H, includes both diagonal and off-diagonal terms for interactions of atomic orbitals.

4.1.1 $H_{\mu\mu}$

Diagonal elements $H_{\mu\mu}$ correspond to the orbital ionization energies of the *s* and *p* orbitals of a given atom.

4.1.2 $H_{\mu\nu}$ (One Center)

Hμv one-center off-diagonal elements:

$$H_{\mu\nu} = U_{\mu}(A) \cdot \delta_{\mu\nu}(\mu, \nu \in A)$$

One-center off-diagonal elements represent the interactions between orbitals on the same atom. The intra-atomic coupling due to orbital hybridization can be non-zero in a tight-bonding model.[7] . However, the MTB/2 method operates in an orthogonal atomic orbital basis, which assumes that orbitals of the same atom are mutually orthogonal. As a result, the one-center off-diagonal terms are set to zero in our orthogonal basis.[7]

4.1.3 $H_{\mu\nu}$ (Two Center)

The two-center off-diagonal element interactions are the core of this model as they describe the coupling between the orbitals located on different atoms. The interactions involve different orbital combinations between s and p, and they account for spatial orientation of the orbitals as well as bond length between atoms.

The two-center off diagonal values rely on the Kolb and Thiel formula

$$H_{\mu\nu} = \pm \beta_{\mu\nu}^{AB}(R_{AB}/a_0)^{1/2}\exp\left(-\lambda_{\mu\nu}^{AB}R_{AB}^2/a_0^2\right)$$

where β represents the strength of the interaction, λ represents the damping factor that ensures interactions decline with increasing distance, R_{AB} is the interatomic distance, and a_0 is the Bohr radius (0.52917 Å).

4.1.4 P-P interactions

P-P interactions can take on sigma (σ) or pi (π) bonding type depending on the relative alignment of the orbital and bond axis. If orbitals are oriented along the same axis (ex: both are px), then the function will check the magnitude of the unit vector component of the bond direction along that axis. If the component is greater than 0.9, the bond is classified as pp_sigma (σ). Otherwise, the bond will be assigned as pp_pi (π).

4.2 Orbital and Repulsion Energy

The total energy of a molecule in the MTB/2 model is the summation of the electronic and repulsion contributions

$$E = \sum_i n_i \varepsilon_i + E_{rep}$$

where the first term represents the electronic energy and the second term represents electron-electron repulsion.[7] The orbital orbital energies, εi, are obtained from diagonalizing the hamiltonian. This calculation relied a generalized eigenvalue problem

$$HC = \varepsilon C$$

where *C* is the molecular orbital as linear combinations of atomic orbitals.

Computed in the occupation_level() and energy_summation() functions, n_i is the orbital occupation number according to the total number of valence electrons.

Following the MTB/2 model, we include the short-range repulsion term E_{rep} to account for the effects of the interatomic repulsion. The summation of the electron repulsion energy is computed over all atomic pairs

$$E_{rep} = \sum_{A>B} G_{AB}$$

where G_{AB} is a two–center potential that takes into account the distance betweens two atoms.[7]

$$G_{AB} = \gamma_{AB} \exp(-\alpha_{AB} R_{AB}) + \omega_{AB} \exp[-6(R_{AB} - r_{AB})^2]$$

The parameters α, γ, ω, r are specific to the atom pair and incorporated from referenced values.[7]
The repulsion term is calculated by iterating over all unique atom pairs.

The summation of the electronic energy and the repulsion energy is then converted from eV to kcal/mol using *f = 23.06054.*

4.3 Heat of Formation

The heat of formation , $\Delta H^0_{f,298}$, was calculated with respected to H_2. Three terms were considered for the heat of formation, with the first being an isolated atomic orbital energy.

$$E_{isol}(A) = n_s U_s(A) + n_p U_p(A)$$

This function estimates the sum of the orbital energies for the isolated atoms that makeup the molecule. This is done by retrieving the number of valence s and p electrons and multiplying them by the corresponding ionization energies U_s and U_p. Doing so accounts for the cost of placing each electron in an isolated atomic orbital. This term allowed for the quantification of energy change that results from bond formation and electron delocalization.

To align with the experimental data, we add the experimental_heat_summ() function which ensures that the calculation of heat of formation accurately reflects the model's prediction of molecular stability and matches the zero reference in the experimental thermochemistry.

In addition to isolation energy and experimental heat sum, the total energy was used to calculate enthalpy [7].

$$\Delta H^0_{f,298} = E - \sum_A E_{isol}(A) + \sum_A H^0_{f,298}(A)$$

The final heat of formation $\Delta H^0_{f,298}$ values are reported in kcal/mol.

4.4 Software Libraries

Within the C++ implementation, we applied several external libraries to support linear algebra operations, file input and output information as well as data parsing. We utilized the Armadillo library for linear algebra and constructing the diagonalized matrices. We also applied the nlohmann/json library to parse the input data in JSON which contained the atomic elements, 3D coordinates and bond orders.

5 Results

Table 1. Experimental vs. predicted energies for selected hydrocarbons (kcal/mol).
The difference reflects the prediction error calculated using (Predicted - Experimental), and the absolute error highlights the magnitude of the deviation. Based on the calculations, hydrocarbons are classified as over or under predicted. All reported energy and enthalpy values are reported with respect to benzene and H_2, respectively.

	Molecule	Experimental	Predicted	Difference	Absolute_Error	Bias
4	pentane	-35.1	214.14300	-249.243	249.24300	Over
3	butane	-30.1	173.63300	-203.733	203.73300	Over
10	trimethylethene	-9.9	179.78700	-189.687	189.68700	Over
9	isobutene	-4.0	168.63100	-172.631	172.63100	Over
2	propane	-25.0	133.21600	-158.216	158.21600	Over
20	benzene	19.7	153.86200	-134.162	134.16200	Over
27	cyclopropane,tris(methylene)	74.8	207.19800	-132.398	132.39800	Over
25	fulvelene	33.8	-58.54170	92.3417	92.34170	Under
12	vinylacetylene	70.4	149.89000	-79.49	79.49000	Over
7	acetylene	54.5	-12.31000	66.81	66.81000	Under
18	cubane	148.7	81.95710	66.7429	66.74290	Under
16	cyclobutene	37.5	-26.56910	64.0691	64.06910	Under
15	cyclopropene	66.2	8.04425	58.15575	58.15575	Under
26	2,4-hexadiyne	70.4	114.87200	-44.472	44.47200	Over
17	fulvene	53.5	95.32050	-41.8205	41.82050	Over
11	diacetylene	111.0	82.01010	28.9899	28.98990	Under
8	propene	6.7	23.08040	-16.3804	16.38040	Over
14	cyclopropane	12.7	28.92750	-16.2275	16.22750	Over
6	ethylene	12.5	20.23880	-7.7388	7.73880	Over
24	benzene	0.0	0.00000	0	0.00000	Under
0	H2	0.0	0.00000	NaN	0.00000	Under

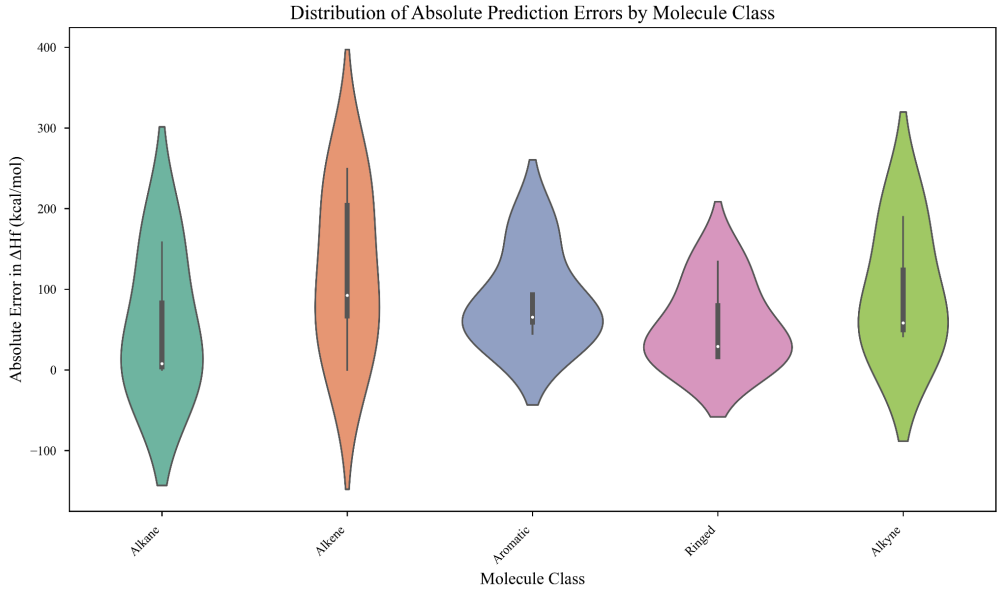

Figure 1. Distribution of absolute prediction errors by molecule class.

The violin plot illustrates the absolute errors for $\Delta H^0_{f,298}$ (kcal/mol) across five molecular classes: alkanes, alkenes, aromatics, cyclic, and alkynes. Within the violin plot is a box plot visualizing the full error distribution, median (white dot), and the interquartile range.

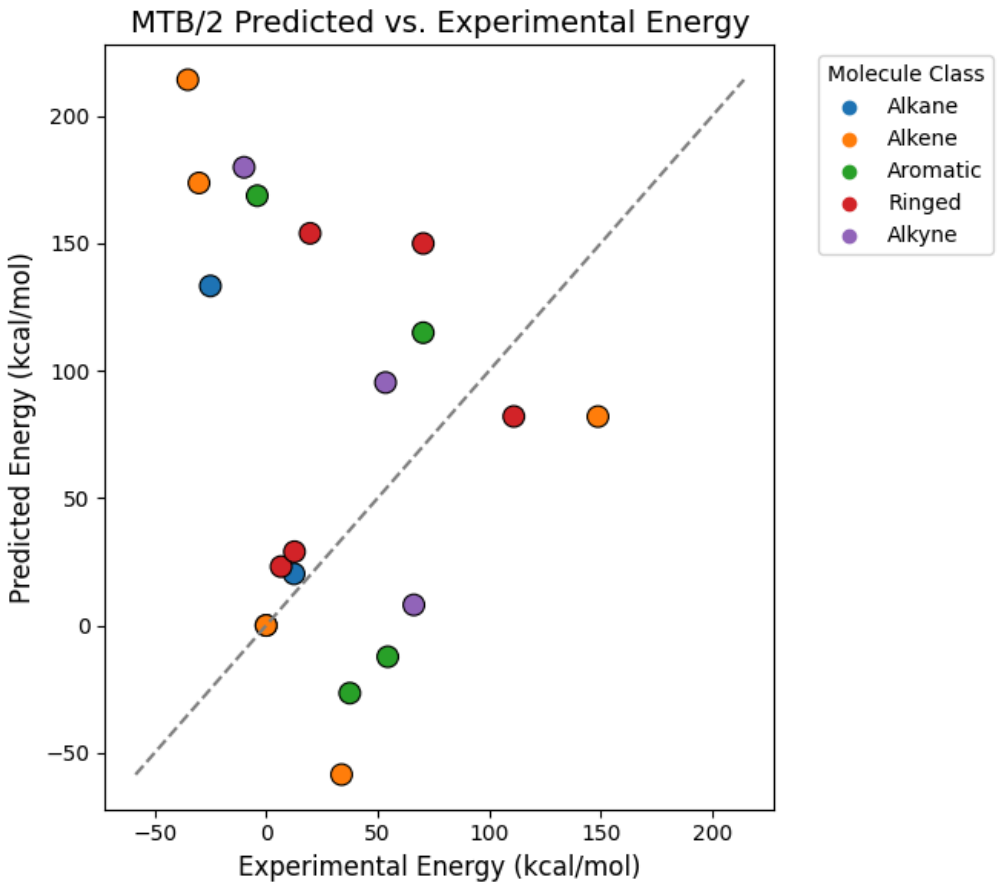

Figure 2. Predicted vs. experimental heat of formation by molecule class. The scatter plot compares the MTB/2 predicted energies against the experimental values in kcal/mol. The colors marked on the scatter plot correspond to the molecule class labeled on the legend. The dashed gray line represents the agreement line in which y = x. Points above the line indicate overprediction and points below indicate under prediction.

5.1 Table 1

After implementing the MTB/2 method, our goal was to evaluate how accurately it predicts heat of formation ($\Delta H^0_{f,298}$) compared to the experimental values. While Voityuk maintained an error deviation of 5 kcal/mol, our implementations showed >100 kcal/mol overpredictions for several systems. Larger over predictions are pronounced in alkanes, branched/substituted alkenes, strained rings with saturation.

In contrast, better performing molecule classes are seen simple alkenes and alkynes. For example, diacetylene, ethylene, and cyclopropene showed absolute errors of 28.99, 7.73, and 58.16 kcal/mol, respectively. Aromatic and cross-conjugated systems also performed better, though some deviations remain.

5.2 Figure 1

In figure 1, the violin plot breaks down the distribution of absolute errors by molecule class. Alkenes show the widest error distribution with some predictions deviating by more than 300 kcal/mol. Alkanes and alkynes also exhibit large areas of distribution. Similarly, figure 2 illustrates that aromatic and ringed compounds display tighter distributions with lowest median error, indicating better consistency and overall accuracy.

5.3 Figure 2

The scatter plot assesses the outcomes between the MTB/2 predicted heats of formation and the experimental values for various hydrocarbon molecule classes. Predictions found above the 1:1 line of agreement indicate that overpredicted heats of formation, while those that deviate below are under predicted. As seen in the figure 1, most predictions deviate above this line, particularly seen in alkenes and alkanes. Aromatic and ringed compounds cluster closer to the line suggesting better agreement between the predicted and reference values for these classes.

Alkanes and alkenes generally lie far above the diagonal line of agreement, suggesting systematic overprediction in σ-rich systems or simpler π systems , whereas aromatics and alkynes cluster closer to the ideal line. This reflects that MTB/2 tends to underestimate the stability of saturated and unsaturated hydrocarbons.

6 Discussion

In an attempt to understand why certain classes of molecules like alkanes and branched alkenes showed large errors, we assess areas where we applied simplifications in the MTB/2 model that could introduce systematic bias.

6.1 Rigid P orbital classification Sigma vs. Pi Threshold

In our model, we classified orbital overlap type in accordance with the MTB//2 based on whether the interatomic vector aligned closely with the orbital axis. Although the classification logic was not specified within the reference, we applied a hard threshold of 0.9 to classify the orbital interactions as either sigma or pi. Although this approach follows the MTB/2 simplicity, it introduces fundamental limitations in which real molecular geometry is rarely perfect. An overly rigid or static orbital threshold allows for misclassification of sigma or pi bonding as it fails to capture any subtle rotations or deformed geometries which is often seen in strained or π-rich systems.[6]

The effects of a misclassified orbital type are rooted in a misassigned β parameter for σ bonding (β_ppσ, -8.420) as it is larger in value than π bonding (β_ppπ, -7.403) (Appendix A). For π-type overlap misclassified as σ, MTB/2 applies an inappropriate strong coupling and inflates the off-diagonal Hamiltonian matrix elements. This results in decreased orbital energies (eigenvalues), and overestimates the stability, ultimately causing errors in the total energy and overly negative heat of formation.

The effects of these misclassifications is evident particularly in alkenes, alkynes and cyclic structures where bond geometries deviate from ideal values.

6.2 Phase Factor

In regards to phase factor, we assumed a constant +1 phase factor for all atomic orbital (AO) overlaps. Essentially treating all interactions between orbitals as constructive, regardless of their orientation in space. While this approach simplifies our MTB/2 model, it does introduce potential error that could be exacerbated by π-rich systems where the sign/phase of the orbital overlap defines whether the interaction is bonding or antibonding. Based on the molecular orbital (MO) theory, when two p orbitals are overlapping in-phase, they form a bonding interaction (π), and when they are out of phase they form an antibonding interaction (π*).[1] By assuming a constant +1 overlap, we ignore this important distinction.

Omitting the phase distinctions within orbital overlap assumes all π overlaps are bonding regardless of antibonding configurations. Since antibonding interactions are never penalized, the system appears more stable than it should be. As a result, the signs and magnitude of the off-diagonal Hamiltonian elements are affected and thereby distort the MO eigenvalues artificially stabilizing the total energy. This downfall of errors explain why we see the worst errors in alkenes and alkynes, where p orbital alignment critically defines bonding character.

6.3 Error Cancellation

6.3.1 Aromatics & Ringed systems

Aromatic and ringed molecules tend to show lower prediction errors, which suggests that MTB/2 handles delocalized π systems better than localized σ systems. This likely comes from the high symmetry and even distribution of π electrons in these molecules where the p orbital shares electron density equally, and the bonds are all very similar.

Even though the MTB/2 model can over- or underestimate individual p–p interactions due to strict σ/π classification or the lack of phase corrections, these consistently small errors across the whole molecule often cancel out. When all the molecular orbital energies are added up, the overestimates and underestimates tend to cancel, giving a total energy that may appear accurate. That's why aromatic systems usually show smaller errors and more reliable $\Delta H^0_{f,298}$ predictions, even with MTB/2's simplified assumptions.

7 Conclusions

Based on Voityuk implementation, MTB/2 method applies a highly efficient and accurate approach to predict the heats of formation for hydrocarbons. This model assumes orthogonality where the overlap matrix is equal to the identity matrix *(S = I),* eliminating the need for an explicit overlap integral and self-consistent (SCF) iterations. This simplification reduces computational complexity in methods like Extended Huckel, CNDO/2, or AM1/PM3.[7]

Our implementation of MTB/2 aligns with the performance trends reported in Voityuk's original work. Particularly, we observe a strong agreement for aromatic systems due to internal error cancellation by delocalized π systems and molecular symmetry. Similarly, branched alkanes and strained molecules showed significant deviations, confirming that MTB/2's lack of a repulsion term and hybridization treatment limits its ability to handle steric effects and non-ideal geometries.

According to Voityuk, MTB/2 highlights potential areas of improvement through parameter tuning, in which our results reveal the deeper structural limitation with rigid σ/π classification and the absence of orbital phase sensitivity. These systematic issues led to overpredictions of π-rich systems and inconsistent behavior in strained rings. As a result, our findings support the MTB/2 baseline method but also demonstrate refinements such as lowering the overlap threshold and applying directional phase logic could enhance its reliability for modern applications.

Overall, our findings support the value of MTB/2 as a baseline semiempirical method. At the same time, they motivate clear directions for refinement that could extend the model's accuracy to a broader range of molecular systems.

8 Future Work

While MTB/2 offers an efficient framework for estimating thermochemical properties of hydrocarbons, our findings highlight several directions for refinement. Our hard threshold of 0.9 for σ/π classification does not account for bond angle distortions or subtle orbital misalignments. Lowering this threshold (to 0.8 or 0.75) could improve $\Delta H^0_{f,298}$ predictions, especially in strained or non-planar systems.[2] As real systems exhibit varying orbital overlaps due to continuous bond angle deformations, and classification based only on a static threshold ignores this directional nuance.

In addition, the model assumes all orbital overlap is constructive (phase factor of +1), neglecting antibonding distribution skews the results in π rich systems.[3] A potential correction involves computing the dot product between the axis and bond vector to account for sign changes in β to reflect the antibonding character.

These combined adjustments can contribute to the enhancement of MTB/2's method by extending accuracy across more chemically complex systems.

References

(1) Ashenhurst, J. (2023, January 8). *Bonding and antibonding pi orbitals*. Master Organic Chemistry.https://www.masterorganicchemistry.com/2017/02/14/molecular-orbital-pi-bond/#notef our
(2) Dewar, M. J., & Thiel, W. (1977). Ground states of molecules. 38. the MNDO method. approximations and parameters. *Journal of the American Chemical Society*, *99*(15), 4899–4907. https://doi.org/10.1021/ja00457a004
(3) Hoffmann, R. (1963). An extended Hückel theory. I. Hydrocarbons. *The Journal of Chemical Physics*, *39*(6), 1397–1412. https://doi.org/10.1063/1.1734456
(4) Kolb, M., & Thiel, W. (1993). Beyond the MNDO model: Methodical considerations and numerical results. *Journal of Computational Chemistry*, *14*(7), 775–789. https://doi.org/10.1002/jcc.540140704
(5) Porezag, D., Frauenheim, Th., Köhler, Th., Seifert, G., & Kaschner, R. (1995). Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Physical Review B*, *51*(19), 12947–12957. https://doi.org/10.1103/physrevb.51.12947
(6) Shi, X., Song, J., & Wei, D. (2025). An analysis method including orbital overlap directions for predicting π electron properties and reactivity vectors. *Nature Communications*, *16*(1). https://doi.org/10.1038/s41467-025-58281-9
(7) Voityuk, A. A. (2008). Thermochemistry of hydrocarbons. back to extended hückel theory. *Journal of Chemical Theory and Computation*, *4*(11), 1877–1885. https://doi.org/10.1021/ct8003222

9 Appendices

Appendix A:[7]

Table 1. Parameters of the MTB/2 Model

Atomic Parameters		
parameter	H	C
U_s^A (eV)	−13.605	−21.559
U_p^A (eV)		−13.507

Bond-Type Parameters			
parameter	H−H	C−H	C−C
λss	0.280	0.275	0.086
λsp		0.218	0.180
$\lambda pp\sigma$			0.186
$\lambda pp\pi$			0.282
βss (eV)	−4.442	−8.574	−5.969
βsp (eV)		−6.813	−6.160
$\beta pp\sigma$ (eV)			−8.420
$\beta pp\pi$ (eV)			−7.403
α_{AB} (Å^{-1})	2.823	2.831	3.401
γ_{AB} (eV)	12.612	99.370	658.659
ω_{AB} (eV)	−0.0791	−0.0340	0.0312
r_{AB} (Å)	2.279	2.843	3.044