# Quantum alchemy beyond singlets: Bonding in diatomic molecules with hydrogen

Emily A. Eikey,[1] Alex M. Maldonado,[2] Charles D. Griego,[2] Guido Falk von Rudorff,[3] and John A. Keith[2, a)]

[1)] *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States*

[2)] *Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States*

[3)] *Faculty of Physics, University of Vienna, Kolingasse 14-16, 1090 Vienna, Austria*

(Dated: 22 November 2021)

Bonding energies are key for the relative stability of molecules in chemical space. Therefore methods employed to search for relevant molecules in chemical space need to capture the bonding behavior for a wide range of molecules, including radicals. In this work, we investigate the ability of quantum alchemy to do so for exploring hypothetical chemical compounds, here diatomic molecules involving hydrogen with various electronic structures. We evaluate equilibrium bond lengths, ionization energies, and electron affinities of these fundamental systems. We compare and contrast how well manual quantum alchemy calculations, i.e. quantum mechanical calculations in which the nuclear charge is altered, and quantum alchemy approximations using a Taylor series expansion can predict these molecular properties. We also investigate the extent of error cancellation of these approaches in terms of ionization energies and electron affinities when using thermodynamic cycles. Our results suggest that the accuracy of Taylor series expansions are greatly improved by error cancellation in thermodynamic cycles, and errors also appear to be generally system-dependent. Taken together, this work provides insights into how quantum alchemy predictions using a Taylor series expansion may be applied to future studies of non-singlet systems as well as which challenges remain open for these cases.

---

[a)] Electronic mail: jakeith@pitt.edu

# I. INTRODUCTION

The ability to efficiently explore chemical and materials space is important for computational screening of promising compounds for new technologies. To make useful, physically relevant computational predictions, hypothetical targets must be physically modeled across chemical and/or materials space, and this usually requires costly self-consistent quantum mechanics (QM) calculations. Considering the vastness of compositional space and the diversity in atom arrangements for any hypothetical composition, it quickly becomes clear that approaches heavily reliant on QM calculations will be computationally expensive and limited in scope. Machine learning models that can interpolate within large QM data sets are one means of overcoming these issues,[1,2] but even more desirable would be reliable protocols that would extrapolate from such data sets and significantly reduce computational burdens.

One such method with promise is quantum alchemy predictions using a Taylor series expansion.[3,4] In short, the Taylor series expansion approximates the so-called alchemical potential energy surfaces (PES) that can be used to conceptualize regions of chemical and materials space. Quantum alchemy approaches have been used to predict energies of various material classes and processes, including molecular systems,[3,5–7] and adsorbate-surface interactions.[8–10] These studies demonstrate a positive outlook for applications of this quantum alchemy methods, but more understanding of the fundamental chemical physics involved in these approximations is still needed before it can be widely adopted. Prior to this work, we evaluated the quantitative accuracy of Taylor series approximations when applied to fundamental energies of atoms in different electronic states and having different charges.[11] That work brought new and fundamental understandings of how and why Taylor series approximations can appear accurate, the significance of using a reference's finite basis set in such predictions, and the importance of exploiting error cancellations when using thermodynamic cycles for energy predictions. The next logical step is to understand the performance of quantum alchemy using a Taylor series expansion on fundamental chemical bonding in molecules.

In this work, we investigated the performance of quantum alchemy using a Taylor series expansion as a function of bond length for simple diatomic molecules involving hydrogen. Specifically, we quantify the accuracy of these methods in predicting equilibrium bond lengths and compare these values to QM-calculated bond length scans and QM-optimized geometries. We then report the ability of manual quantum alchemy and quantum alchemy using a Taylor series expansion to predict relative minimum energies through ionization energy and electron affinity predictions from thermodynamic cycles. These results indicate that the Taylor series introduces some additional errors beyond self-consistent, manual quantum alchemy calculations, as expected, but the magnitude of these errors both appear to be system-dependent and are significantly reduced when employing judiciously constructed thermodynamic cycles. Overall, this work reports how well manual quantum alchemy and quantum alchemy using a Taylor series expansion predict equilibrium bond lengths and bond energies as well as molecular properties based on electronic structures such as ionization potentials and electron affinities of relatively simple and fundamental chemically bonded systems.

## II. METHODS

### A. QM calculations and geometry optimizations

QM calculations were performed with the open-source Gaussian basis set code, PySCF (v1.7.6)[12–14]. Coupled cluster calculations with single, double, and perturbative triples based on restricted and unrestricted (when necessary) Hartree–Fock (HF) references used default convergence criteria. We modeled XH diatomic molecules from LiH to NeH with charges of 0, 1±, and 2+ as denoted in Figure 1. QM calculations were performed using the cc-pV5Z basis set[15–19]. Geometry optimizations employed geomeTRIC[20] (v0.9.7.2) as implemented in PySCF with default convergence criteria. When possible, we compared QM calculations to experimental quantities including first ionization energies[21,22] and electron affinities[22]. All QM calculations used for quantum alchemy predictions using a Taylor series expansion converged, but some QM calculations using integer-valued differences in nuclear charge (manual quantum alchemy calculations, see below) did not converge, and these are discussed in the SI (see Table S1). The non-converged calculations do not influence conclusions drawn from the data.

### B. Equilibrium bond lengths and energies

Bond length scans ranged from 0.6 Å to 1.9 Å using a step size of 0.1 Å. To identify more-refined equilibrium bond lengths and minimum energies of each diatomic molecule, we selected the minimum energy point from the scan and the two closest preceding and succeeding points and fit those five points to a fourth order polynomial to best locate the interpolated minimum.

### C. Manual quantum alchemy predictions using modified QM calculations

Alchemical potential energy surfaces (PESs) were calculated at each point length along the bond length scan. Manual quantum alchemy predictions were obtained by starting from a reference diatomic molecule XH, modifying the nuclear charge ($Z$) of atom X, and calculating the energy of the new diatomic with hydrogen using QM. For example, starting from the $^2\Pi$ state of OH as a reference, an analogous $^2\Pi$ state for $NH^-$ can be obtained by decreasing the nuclear charge on the O atom by one ($\Delta Z = -1$). Alternatively, OH could be used as a reference for the $^2\Pi$ of $FH^+$, which represents a $\Delta Z = +1$ transmutation. Manual quantum alchemy calculations use the basis sets of the reference diatomic to calculate the energy of a target diatomic, where the reference and target only differ in terms of the nuclear charge of atom X. All alchemical transmutations used here will retain the same number of electrons as the reference system. Subsequent interpolations as described above allowed us to determine equilibrium bond lengths and bond energies for all target molecules (see Figure S1).

### D. Quantum alchemy predictions using a Taylor series expansion

For more specific details about our approach, we refer to our earlier manuscript.[11] In short, a Taylor series expansion around the energy of a reference system with respect to an arbitrary alchemical pathway is used to estimate the energy of a target system. Although other mathematically equivalent quantum alchemy schemes exist,[5,9,23–26] we employ a central finite differences approach to calculate terms in the Taylor series.[26] Using the finite differences method, various orders of the Taylor series expansion can be determined. Thus, either manual quantum alchemy or quantum alchemy using a Taylor series expansion may be used to approximate alchemical PESs defined at each scanned bond length, which in turn provides approximate bond length scans for target molecules. The resulting bond length scans can be fit using the procedure described above to determine equilibrium bond lengths and bond energies that can be compared to reference QM data.

### E. Thermodynamic cycles to predict diatomic properties

Thermodynamic cycles were used to calculate first ionization energies and electron affinities for diatomic molecules. We refrain from predicting properties of systems with a 2+ charge state (e.g., second ionization energies), as these did not show distinct minima within the range of distances studied in this work. An example thermodynamic cycle is shown in Figure S2. The top leg (state A to B) represents a reference property ($\Delta E_1$), and the bottom leg (state C to D) represents the target property ($\Delta E_4$):

$$0 = \Delta E_4 - \Delta E_1 - \Delta E_3 + \Delta E_2 \tag{1}$$

$$\Delta E_4 = \Delta E_1 + \Delta E_3 - \Delta E_2 \tag{2}$$

QM calculations are used to calculate $\Delta E_1$. To calculate $\Delta E_4$, $\Delta \mathrm{E}_2$ and $\Delta \mathrm{E}_3$ can be obtained using manual quantum alchemy calculations (obtained with self-consistent field calculations using different nuclear charges as described above) or quantum alchemy using a Taylor series expansion. Since the vertical legs are predicted using one of these quantum alchemy schemes, the states involved in each vertical leg must contain the same number of electrons (i.e., states A and C and states B and D). Within the scope of this work, several reference and target combinations exist. For instance, the first ionization energy of OH (OH $\rightarrow$ $\mathrm{OH}^+ + e^-$, $\Delta E_4$) can be predicted using the following references for $\Delta E_1$: $\mathrm{NH}^- \rightarrow \mathrm{NH} + e^-$ and $\mathrm{FH}^+ \rightarrow \mathrm{FH}^{2+} + e^-$, using transmutations of $\Delta \mathrm{Z} = +1$ and -1, respectively. A goal of this work was to gain insight into which simple diatomic references would be the most accurate and why.

## III. RESULTS AND DISCUSSION

### A. QM calculations and comparisons to experiment

We first compared conventional QM-calculated bond curves, ionization energies, and electron affinities of the diatomic molecules and compared these results to experiment.[21,22] Equilibrium bond lengths were determined by optimizing the geometry of each dimer and by fitting the calculated bond length scan using the procedure described above. Table I reports mean absolute error (MAE), root mean square error (RMSE), and maximum error (max error) for equilibrium bond lengths using both methods. Since the level of QM we used was relatively high, the calculated data were generally quite accurate. As both methods provided reasonable results, the fitted QM data shown here was used to compare manual quantum alchemy calculations and quantum alchemy predictions using a Taylor series expansion, which were also determined using the fitting procedure.

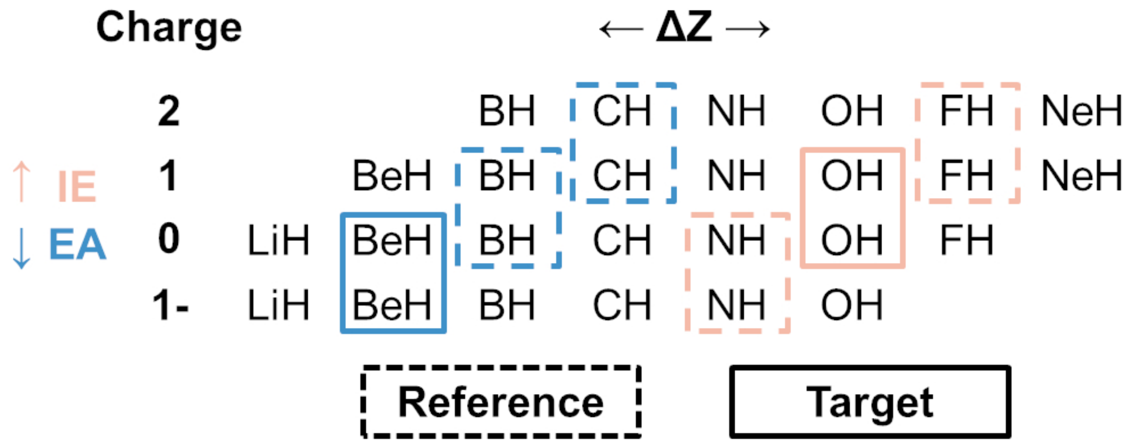

FIG. 1. Illustration of the chemical space sampled in this work. Diatomic molecules from LiH to NeH were studied with different charge states, as labeled on the left. Ionization energies (IE) and electron affinities (EA) were calculated using QM and predicted using quantum alchemy schemes using thermodynamic cycles. Examples of target states (solid boxes) and references for each of those targets (dotted boxes) are shown. Quantum alchemy requires that the number of electrons remain constant between the reference and targets, which limit the reference and target combinations.

Ionization energy describes the energy required to remove an electron from a neutral diatomic molecule to form a cationic molecule:

$$\mathrm{XH} \rightarrow \mathrm{XH}^{+} + e^{-} \tag{3}$$

Electron affinity describes the change in energy to add an electron to a neutral system:

$$\mathrm{XH} + e^- \rightarrow \mathrm{XH}^- \tag{4}$$

In both of these schemes, the energies and equilibrium bond lengths of the neutral and charged diatomic molecules were taken from interpolative fitting procedures described above. Equilibrium bond lengths and system energies from the interpolations were used to calculate ionization energies and electron affinities that were then compared to experiment.[21,22] The MAE for ionization energies and electron affinities were determined to be 0.094 eV and 0.264 eV, respectively (Tables I and S3). As expected, electron affinity calculations were somewhat less accurate than ionization potential calculations using the cc-pV5Z basis sets. Augmented functions are typically prescribed for treating anionic systems, but we found these basis sets difficult to reach converged results for all systems, and thus we chose the smaller cc-pV5Z basis set that allowed us to do consistent reference data calculations across all diatomic systems at different bond lengths. These data were used to compare quantum alchemy predictions to, and the QM calculations used to obtain these data were used as reference calculations for quantum alchemy predictions using a Taylor series of other systems (see Figure 1).

| Bond Length | MAE (Å) | RMSE (Å) | Max Error (Å) |
|---|---|---|---|
| Geometry-optimized | 0.007 | 0.007 | 0.011 |
| Fit | 0.007 | 0.008 | 0.012 |

| Property | MAE (eV) | RMSE (eV) | Max Error (eV) |
|---|---|---|---|
| IE | 0.094 | 0.161 | 0.378 |
| EA | 0.264 | 0.280 | 0.361 |

TABLE I. Summary of MAE, RMSE, and maximum error for QM-calculated equilibrium bond lengths, ionization energies, and electron affinities compared to experiment.

## B. Bond length scan predictions using quantum alchemy schemes

To simplify alchemical modeling as much as possible, it is customary to carry out atom transmutations using the same interatomic coordinates and the same number of electrons as the reference calculation. However, transmutations in this way on a molecular structure will invariably result in modeling a compound in a non-equilibrium structure. To assess how accurately alchemical methods can predict actual equilibrium structures, we used the sequence of QM calculations from a bond length scan as a sequence of references used for alchemical transmutations based on two quantum alchemy schemes, as described below. In other words, alchemical PESs were calculated at each bond length of the reference scan (see Figure S1).

First, we can consider using what we refer to as manual quantum alchemy, which refers to using a reference's basis sets, modifying the nuclear charge of the heteroatom with the diatomic, and running a QM calculation, to calculate PESs at each bond length. For example, consider the alchemical PES for the neutral NH diatomic molecule that contains eight electrons in the $^3\Pi$ configuration. Within the scope of this work, changing the nuclear charge of N can result in either $CH^-$, $OH^+$, or $FH^{2+}$. Alternatively, one can also use any of these species as a reference case to make any other PES for the same targets at different bond lengths. After all PESs are constructed at each bond length, the equilibrium bond length and minimum energy can be calculated for each PES by fitting the bond length scan. Figure 2 illustrates the quantum alchemy-calculated bond length scans and the QM-calculated target bond length scan each attempts to predict.

Next, we can consider predicting bond length scans via quantum alchemy using a Taylor series expansion. To do so, alchemical PESs were approximated using a Taylor series expansion about each reference at each bond length. For every target of interest, the corresponding points on each alchemical PESs across the bond length scan can be used to construct a predicted bond length scan. Since different orders of the Taylor series can be truncated, we can also investigate how various orders impact the accuracy of the predicted bond length scans. If we again use the example above with eight electron diatomic molecules, we can consider each system as reference and target and assess how well quantum alchemy using a Taylor series expansion predicts bond length scans (Figure 2). Manual quantum alchemy-calculated and Taylor series-predicted bond length scans can be plotted for all other reference/target diatomic molecules, which are freely available at github.com/keithgroup/qa-atoms-dimers.

For manual quantum alchemy calculated bond length scans, we observe deviations from the QM-calculated scans, indicating that there is an intrinsic alchemical error associated with using a reference's basis set to model another system. Previous works, including our own,

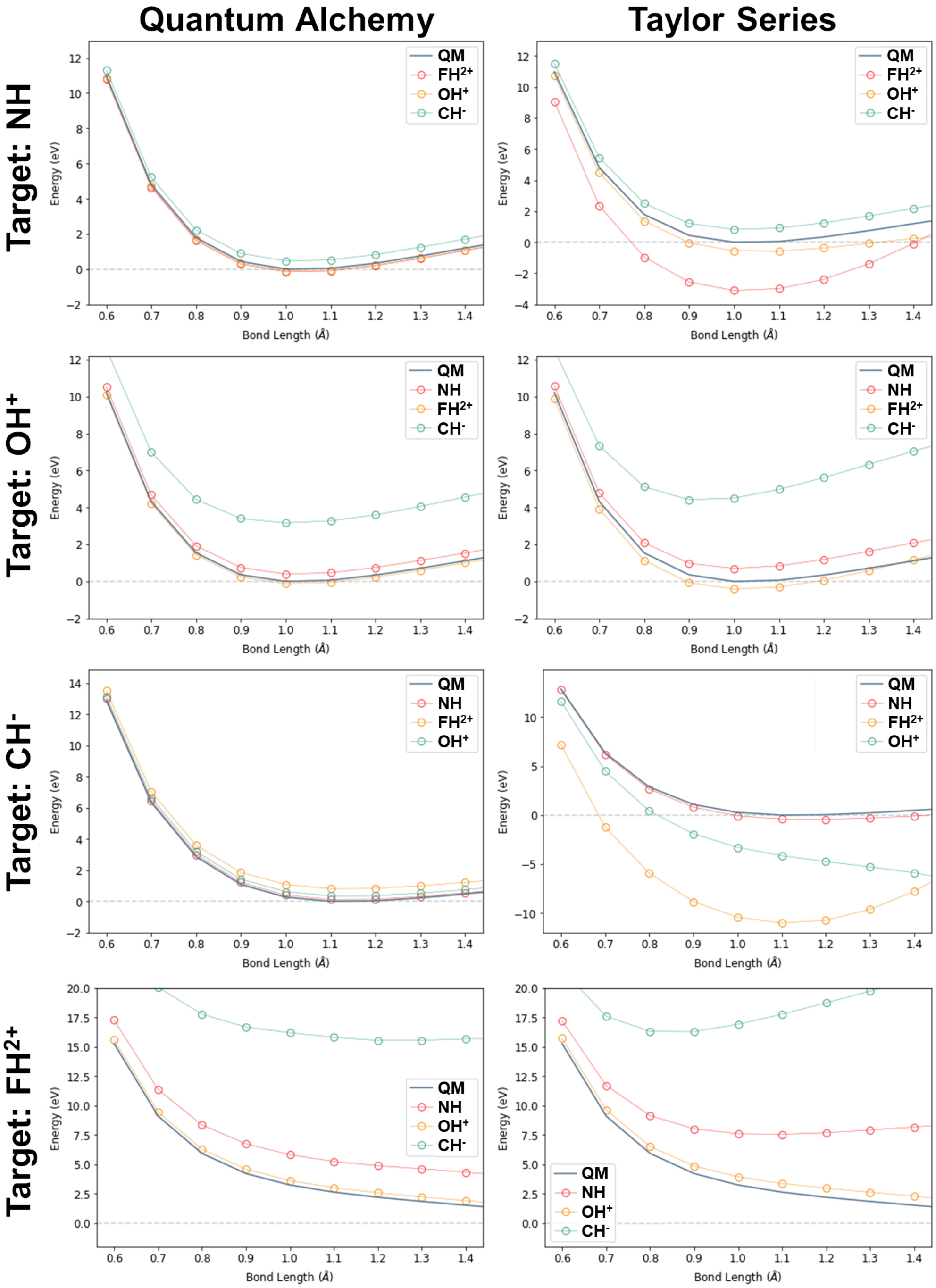

FIG. 2. Bond length scans for eight electron systems using manual quantum alchemy calculations and quantum alchemy using a second order Taylor series expansion. Plots are truncated to 1.4 Å, since some deviations stemming from QM calculations at higher bond lengths are observed (see SI for details). Reference systems are reported in the legends of each plot, and target systems are reported on the left.

have touched on this potential issue.[5,11] Quantum alchemy-calculated bond length scans for all reference and target combinations were fit to determine the predicted equilibrium bond length, and these errors with respect to QM-fitted bond length scans are reported in Table II. We exclude predicting bond length scans for targets that did not have minima as calculated by bond length scans and/or geometry optimizations (see SI for more details). Quantitatively, we find that while deviations are observed in these calculated bond length scans, overall, the equilibrium bond lengths are reasonably accurate (MAE $\sim$ 0.015 Å). We also assessed the impact of predicting bond length scans using a Taylor series expansion. We find that generally, Taylor series truncated after the second order provide the most accurate bond length scan predictions (Table S2).

Overall, smaller errors in equilibrium bond lengths are observed for manual quantum alchemy than quantum alchemy using a Taylor series. However, approximating alchemical PESs at each bond length as a Taylor series expansion is expected to introduce some error. Errors in equilibrium bond length predictions can be discussed as a function of $\Delta$Z, or the change in nuclear charge from the reference to the target diatomic. For both manual quantum alchemy and quantum alchemy using a Taylor series, predicted equilibrium bond lengths are most accurate for $\Delta\mathrm{Z} = \pm 1$. In quantum alchemy, the basis sets of the reference molecule are used to model the target molecules, and so basis sets will be most appropriate for smaller changes in nuclear charge than larger ones. Similarly, the Taylor series approximates the alchemical PESs and resulting bond length scan by using information about the reference system. At this order, the larger the changes in nuclear charge from the reference, the less local predictions become and the larger deviations will be from the reference system.

While no clear trend is observed in manual quantum alchemy MAEs for equilibrium bond lengths as function of the sign of $\Delta$Z, MAEs appear to be larger for positive $\Delta$Z values than negative ones for quantum alchemy predictions using a Taylor series. Therefore, this data suggests that target equilibrium bond lengths are more accurately predicted with positive nuclear charge perturbations ($\Delta Z > 0$). These results provide important insights into how and when quantum alchemy using a Taylor series expansion may be used to predict bond length scans and determine equilibrium bond lengths. However, to further investigate the applicability of this method to aid in geometry optimizations, it is important to assess the accuracy of the relative energies at those equilibrium bond lengths.

| Method | ΔZ | MAE (Å) | RMSE (Å) | Max Error (Å) |
|---|---|---|---|---|
| Quantum Alchemy | -3 | 0.018 | 0.027 | 0.049 |
|  | -2 | 0.017 | 0.024 | 0.047 |
|  | -1 | 0.012 | 0.020 | 0.049 |
|  | 1 | 0.007 | 0.012 | 0.034 |
|  | 2 | 0.035 | 0.040 | 0.060 |
|  | Overall | 0.015 | 0.023 | 0.060 |
| Taylor Series | -3 | 0.091 | 0.117 | 0.217 |
|  | -2 | 0.277 | 0.374 | 0.698 |
|  | -1 | 0.031 | 0.057 | 0.217 |
|  | 1 | 0.021 | 0.026 | 0.058 |
|  | 2 | 0.109 | 0.119 | 0.193 |
|  | Overall | 0.097 | 0.189 | 0.698 |

TABLE II. Summary of MAE, RMSE, and maximum error in equilibrium bond lengths as calculated using manual quantum alchemy or predicted using a Taylor series with respect to equilibrium bond lengths determined using QM-fitted bond length scans as a function of ΔZ.

### C. Diatomic property predictions using thermodynamic cycles

To assess the geometry response of diatomic molecules to transmutations using quantum alchemy, not only must we assess the accuracy of bond length scan and equilibrium bond length predictions, but we must also understand how well relative equilibrium energies are predicted. To this end, we have chosen to predict ionization energies and electronic affinities for diatomic molecules containing hydrogen, as these properties have been robustly measured.

To calculate ionization energies and electron affinities using QM and to predict these properties using quantum alchemy, we employ thermodynamic cycles (Figure S2). Properties calculated using this scheme from QM-calculated values will again serve as a reference for quantum alchemy calculations and predictions.

For example, consider a thermodynamic cycle that can be used to prediction the ionization energy of $NH \rightarrow NH^{+} + e^{-}$. $OH^{+} \rightarrow OH^{2+} + e^{-}$ could be used as a reference process. In this example, $\Delta E_1$ represents the energy difference for $OH^{+} \rightarrow OH^{2+} + e^{-}$. We expect errors will arise when using a finite basis set for O in $OH^{+}$ to predict the energy of NH ($\Delta E_2$) as well as that in $OH^{2+}$ to predict $NH^{+}$ ($\Delta E_3$), but we have recently shown that these errors, although still present, are significantly reduced when a thermodynamic cycle is employed.[11] In order to determine the energies used in the calculation of $\Delta E_1$, bond length scans of $OH^{+}$ and $OH^{2+}$ were fit separately and the minimum energies were calculated. To predict the energies of NH and $NH^{+}$, quantum alchemy or Taylor series predictions were made at each bond length using $OH^{+}$ and $OH^{2+}$ as references, respectively. More specifically, these energies were predicted by either altering the nuclear charge of O by $-1$ to model N or by using a Taylor series expansion about these references. The resulting new bond length scans were again fit to determine the minimum energies. Using Equation 2, the ionization energy could be calculated.

To gain a better understanding of how well manual quantum alchemy and quantum alchemy using a Taylor series expansion predict these diatomic properties, we calculated MAE, RMSE, and maximum errors for predicted ionization energies and electron affinities using these two quantum alchemy schemes (Tables III, S4, and S5). We find that the Taylor series truncated after the second order provides the most accurate predictions (see Tables S4 and S5 for other Taylor series orders), for some cases reaching DFT accuracy. Higher order predictions were recently shown to require high numerical precision and tight convergence criteria,[26] suggesting that in this work, orders higher than the second may require derivatives to be calculated with much higher precision that used here. When greater numerical precision is used, we expect errors in Taylor series predictions would be decreased further and approach the reference QM accuracies. Overall and as expected, manual quantum alchemy provides more accurate predictions when comparing individual properties.

To better understand the trends in alchemical errors using thermodynamic cycles to predict these diatomic properties, we can calculate the error statistics for each $\Delta$Z. This may be most easily illustrated in Figure 1, where the change in nuclear charge from reference to target systems is different for ionization energies and electron affinities. Table III reports the MAE, RMSE, and maximum errors for $\Delta Z = \pm 1$ for ionization energies and $\Delta Z =$ -1 and -2 for electron affinities for both manual quantum alchemy and quantum alchemy using a Taylor series expansion. Even when errors are broken down by $\Delta$Z, the second order

| Method | Property | ΔZ | MAE (eV) | RMSE (eV) | Max Error (eV) |
|---|---|---|---|---|---|
| Quantum Alchemy | IE | -1 | 0.019 | 0.021 | 0.030 |
| | | 1 | 0.006 | 0.007 | 0.011 |
| | | Overall | 0.012 | 0.016 | 0.030 |
| | EA | -1 | 0.227 | 0.228 | 0.252 |
| | | -2 | 0.581 | 0.584 | 0.698 |
| | | Overall | 0.404 | 0.443 | 0.698 |
| Taylor Series | IE | -1 | 0.095 | 0.119 | 0.228 |
| | | 1 | 0.251 | 0.270 | 0.448 |
| | | Overall | 0.173 | 0.209 | 0.448 |
| | EA | -1 | 0.227 | 0.252 | 0.351 |
| | | -2 | 2.819 | 3.115 | 4.621 |
| | | Overall | 1.523 | 2.210 | 4.621 |
| PBE/def2-QZVP | IE | Overall | 0.110 | 0.143 | 0.201 |
| PBE/def2-QZVP | EA | Overall | 0.241 | 0.251 | 0.329 |

TABLE III. Summary of MAE, RMSE, and maximum error for ionization energies and electron affinities calculated using quantum alchemy or a Taylor series expansion using thermodynamic cycles with respect to QM as a function of ΔZ. DFT comparison values for PBE/def2-QZVP are shown for the subset of systems included in the GMTKN55 data set[27].

Taylor series expansion remains the most accurate (Table S5). As expected, ΔZ = -1 MAEs are smaller than ΔZ = -2 for electron affinities for both quantum alchemy schemes. These observations are in line with what we observed above in trends of equilibrium bond length predictions, where predictions should be most accurate for target diatomic molecules that are closer to the reference diatomic in terms of nuclear charge.

While manual quantum alchemy errors are smaller for ΔZ = +1 than ΔZ = -1, this is opposite of what is observed for quantum alchemy using a Taylor series. Previous work has suggested that the asymmetry in error is a result of using the basis sets of a reference to describe target systems.[5] Our results allow us to directly test the impact of the direction of nuclear charge changes. This work shows that errors in Taylor series predictions are lowest when the nuclear charge of the reference is decreased to reach a desired target, suggesting that the basis set of the reference may be more appropriate to model systems with a lower nuclear charge. These dependencies on basis set may also explain why errors for ionization energies are smaller than those for electron affinities for the same $\Delta \mathrm{Z}$ values. If this hypothesis holds, we anticipate that saturated basis sets and/or approaching the complete basis set limit will lead to a less noticeable difference in errors for $\Delta \mathrm{Z} = \pm 1$ predicted using a Taylor series expansion.

However, we find very similar MAEs between quantum alchemy and Taylor series electron affinity predictions using $\Delta Z = -1$. Examining the second order Taylor series predictions on a case-by-case basis reveals that the second-order Taylor series expansion generally induces errors that sometimes improves or worsens electron affinity predictions—effectively cancelling out in the MAE giving the facade of nonexistent errors. Table S6 reports second order Taylor series and manual quantum alchemy errors for electron affinities with respect to QM for $\Delta \mathrm{Z}$ = -1; for example, using a Taylor series, $\mathrm{NH}^+ \rightarrow \mathrm{CH}$, $\mathrm{OH}^+ \rightarrow \mathrm{NH}$, and $\mathrm{FH}^+ \rightarrow \mathrm{OH}$ (for simplicity, only neutral target predictions are shown) provides more accurate predictions than manual quantum alchemy. These results suggest that negative nuclear charge perturbations ($\Delta Z < 0$) with second order Taylor series predictions can lead to fortuitous error cancellation in electron affinity predictions.

When considering the accuracy of ionization energy and electron affinity predictions, the accuracy of bond length predictions must also be considered. Although equilibrium bond lengths were more accurately predicted for $\Delta \mathrm{Z}$ = 1 than -1 for both quantum alchemy schemes, these trends do not hold in all cases for ionization energy predictions. Quantum alchemy predictions follow this trend, but for second order Taylor series approximations, ionization energy predictions are more accurate for $\Delta \mathrm{Z}$ = -1. These results suggest that even while equilibrium bond length predictions may be less accurate for Taylor series predictions when $\Delta \mathrm{Z}$ = -1, the relative energies predicted using the thermodynamic cycle are actually more accurate for $\Delta \mathrm{Z}$ = -1. Therefore, the accuracy of quantum alchemy using a Taylor series expansion may be dependent on the desired application.

## IV. CONCLUSIONS

In all, we have evaluated the performance of manual quantum alchemy and quantum alchemy using a Taylor series expansion for predicting bond length scans, ionization energies, and electron affinities of diatomic molecules with hydrogen. Bond length scans were predicted by generating alchemical PESs at each bond length using both quantum alchemy schemes. Using the bond length scans, equilibrium bond lengths and energies were determined by minimizing a polynomial fit. Comparing predictions to CCSD(T) calculations, we find that equilibrium bond lengths can be predicted within an MAE of 0.015 Å using manual quantum alchemy or within an MAE of 0.097 Å using second order Taylor series approximations. In diatomic property predictions, we again find that quantum alchemy provides more accurate predictions than second order Taylor series approximations. We can gain more understanding of the fundamental factors that impact the accuracy of these methods by comparing their performance as a function of the change in nuclear charge from the reference diatomic to the target diatomic. As expected, smaller changes in nuclear charge lead to more accurate predictions. For equilibrium bond length predictions, increasing the nuclear charge of the reference leads to more accurate predictions than when the nuclear charge is decreased. However, these trends do not always hold for relative energy predictions. These results suggest that the factors influencing second order Taylor series accuracy may be application-dependent. This work provides fundamental insight into the use of manual quantum alchemy and quantum alchemy using a Taylor series for bond length scan and relative energy predictions. Although errors are larger for second order Taylor series predictions, we anticipate that both equilibrium bond length and relative energy predictions that these errors will significantly improve when higher order Taylor series predictions are calculated accurately with higher numerical precision, providing a positive outlook for quantum alchemy to be used in geometry optimizations and property predictions.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in GitHub at github.com/keithgroup, repositories qa-dimers-data and qa-atoms-dimers.

## ACKNOWLEDGMENTS

E.A.E. acknowledges support from the Hurd W. Safford Graduate Teaching Fellowship and the Kenneth P. Dietrich School of Arts & Sciences Graduate Fellowship. We acknowledge financial support from the U.S. National Science Foundation: GRFP-1747452 (for C.D.G.) and CHE-1856460 (J.A.K.) and the R. K. Mellon foundation and the University of Pittsburgh's Swanson School of Engineering. Calculations were run using resources provided by the University of Pittsburgh Center for Research Computing.

## REFERENCES

[1] J. E. Saal, A. O. Oliynyk, and B. Meredig, "Machine learning in materials discovery: Confirmed predictions and their underlying approaches," Annu. Rev. Mater. Res. **50**, 49–69 (2020).

[2] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, "Combining machine learning and computational chemistry for predictive insights into chemical systems," Chem. Rev. **121**, 9816–9872 (2021).

[3] G. F. von Rudorff and O. A. von Lilienfeld, "Alchemical perturbation density functional theory," Phys. Rev. Res. **2**, 023220 (2020).

[4] O. A. von Lilienfeld, "Accurate ab initio energy gradients in chemical compound space," J. Chem. Phys. **131**, 164102 (2009).

[5] G. Domenichini, G. F. von Rudorff, and O. A. von Lilienfeld, "Effects of perturbation order and basis set on alchemical predictions," J. Chem. Phys. **153**, 144118 (2020).

[6] S. Fias, K. Y. S. Chang, and O. A. Von Lilienfeld, "Alchemical normal modes unify chemical space," J. Phys. Chem. Lett. **10**, 30–39 (2019).

[7] G. F. von Rudorff and O. A. von Lilienfeld, "Rapid and accurate molecular deprotonation energies from quantum alchemy," Phys. Chem. Chem. Phys. **22**, 10519–10525 (2020).

[8] C. D. Griego, K. Saravanan, and J. A. Keith, "Benchmarking computational alchemy for carbide, nitride, and oxide catalysts," Adv. Theory Simul. **2**, 1800142 (2019).

[9] C. D. Griego, J. R. Kitchin, and J. A. Keith, "Acceleration of catalyst discovery with easy, fast, and reproducible computational alchemy," Int. J. Quantum Chem. **121**, e26380 (2020).

[10] D. Sheppard, G. Henkelman, and O. A. von Lilienfeld, "Alchemical derivatives of reaction energetics," J. Chem. Phys. **133**, 084104 (2010).

[11] E. A. Eikey, A. M. Maldonado, C. D. Griego, G. F. von Rudorff, and J. A. Keith, "Evaluating quantum alchemy of atoms with thermodynamic cycles: Beyond ground electronic states," arXiv preprint (2021).

[12] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, *et al.*, "Recent developments in the PySCF program package," J. Chem. Phys. **153**, 024109 (2020).

[13] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, "Pyscf: the python-based simulations of chemistry framework," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1340 (2018).

[14] Q. Sun, "Libcint: An efficient general integral library for g aussian basis functions," J. Comp. Chem. **36**, 1664–1671 (2015).

[15] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning, and A. K. Wilson, "Gaussian basis sets for use in correlated molecular calculations. vii. valence, core-valence, and scalar relativistic basis sets for li, be, na, and mg," Theor. Chem. Acc. **128**, 69–82 (2011).

[16] D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. iv. calculation of static electrical response properties," J. Chem. Phys. **100**, 2975–2988 (1994).

[17] D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. iii. the atoms aluminum through argon," J. Chem. Phys. **98**, 1358–1371 (1993).

[18] R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," J. Chem. Phys. **96**, 6796–6806 (1992).

[19] T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007–1023 (1989).

[20] L.-P. Wang and C. Song, "Geometry optimization made simple with translation and rotation coordinates," J. Chem. Phys. **144**, 214108 (2016).

[21] A. Kramida, Y. Ralchenko, J. Reader, and and NIST ASD Team, NIST Atomic Spectra Database (ver. 5.8), [Online]. Available: `https://physics.nist.gov/asd` [2021, October 20]. National Institute of Standards and Technology, Gaithersburg, MD. (2020).

[22] W. M. Haynes, D. R. Lide, and T. J. Bruno, *CRC Handbook of Chemistry and Physics*, edited by W. M. Haynes, Vol. 2016–2017, 97th edition (CRC Press, 2017).

[23] R. Balawender, M. A. Welearegay, M. Lesiuk, F. De Proft, and P. Geerlings, "Exploring chemical space with the alchemical derivatives," J. Chem. Theory Comput. **9**, 5327–5340 (2013).

[24] R. Balawender, M. Lesiuk, F. De Proft, and P. Geerlings, "Exploring chemical space with alchemical derivatives: Bn-simultaneous substitution patterns in $c_60$," J. Chem. Theory Comput. **14**, 1154–1168 (2018).

[25] R. Balawender, M. Lesiuk, F. D. Proft, C. V. Alsenoy, and P. Geerlings, "Exploring chemical space with alchemical derivatives: Alchemical transformations of h through ar and their ions as a proof of concept," Phys. Chem. Chem. Phys. **21**, 23865–23879 (2019).

[26] G. F. von Rudorff, "Arbitrarily accurate quantum alchemy," J. Chem. Phys. (2021), 10.1063/5.0073941.

[27] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, "A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions," Phys. Chem. Chem. Phys. **19**, 32184–32215 (2017).