From quantum alchemy to Hammett’s equation: Revisiting the covalent bond

Michael J. Sahre,1,2 Guido Falk von Rudorff,1 and O. Anatole von Lilienfeld3,4,5

1University of Vienna, Faculty of Physics, Kolingasse 14-16, 1090 Vienna, Austria
2University of Vienna, Vienna Doctoral School in Chemistry (DoSChem), Währinger Str. 42, 1090 Vienna, Austria.
3Vector Institute for Artificial Intelligence, Toronto, ON, M5S 1M1, Canada
4Departments of Chemistry, Materials Science and Engineering, and Physics, University of Toronto, St. George Campus, Toronto, ON, Canada
5Machine Learning Group, Technische Universität Berlin and Institute for the Foundations of Learning and Data, 10587 Berlin, Germany

We present an intuitive and general analytical approximation estimating the energy of covalent single and double bonds between participating atoms in terms of their respective nuclear charges with just three parameters, \( E_{AB} \approx a - bZ_AZ_B + c(Z_A^{7/3} + Z_B^{7/3}) \). The functional form of our expression models an alchemical atomic energy decomposition between participating atoms A and B. After calibration, reasonably accurate bond energy estimates are obtained for hydrogen-saturated diatomics composed of p-block elements coming from the same row 2 \( \leq n \leq 4 \) in the periodic table. Corresponding changes in bond energies due to substitution of atom B by C can be obtained via simple formulas. While being of different functional form and origin, our model is as simple and accurate as Pauling’s well-known electronegativity model. Analysis indicates that the model’s response in covalent bonding to variation in nuclear charge is near-linear—which is consistent with Hammett’s equation.

I. INTRODUCTION

Due to their direct link to thermodynamics stability, computational predictions of binding trends among molecules and materials have greatly improved design choices in the chemical sciences. Unfortunately, while numerical bonding estimates obtained from modern and computationally demanding quantum methods are very accurate and predictive, they are at the same time difficult to grasp with human intuition due to the inherent complexity of their solutions to the electronic Schrödinger equation. Already very approximate models, however, can yield intuitive descriptions of important bonding features. For example, Lewis’ simple concept of binding electron pairs, in conjunction with the Aufbau principle, imposes relevant constraints from Pauli’s exclusion principle or the delocalization of \( \pi \)-electrons can be easily understood within Huckel theory (FIG 1, left). Pauling’s model for covalent bond energies is based on the idea to decompose the wavefunction within valence bond theory into ionic and covalent terms, leading to the chemically intuitive concept of electronegativity (FIG 1, center). Because of their universality due to their foundation in quantum mechanics, these basic models are applicable across chemical compound space, and have proven powerful for advancing our understanding of chemistry—despite their approximate nature.

Less approximate and more recently, energy decomposition methods have been introduced to provide a more detailed understanding of chemical bonding. Some methods decompose the energy into different physical contributions while others partition it onto atoms. Partitioning of quantum mechanical observables onto the constituting parts of the quantum many-body system can be done in arbitrarily many ways. Also within the framework of computational alchemical perturbation density functional theory one can meaningfully quantify the effect of different binding partners on the atomic energy of every atom in the system. Here, we introduce a model of covalent bonding which we compare to molecular orbital (MO) and valence bond (VB) theory in FIG 1. The model directly emerges from quantum alchemy (QA) based atomic energy decomposition arguments. It is, to the best of our knowledge new, yet as simple and accurate as Pauling’s electronegativity model and of a distinctly different functional form and origin. We thus believe that it enables a fresh and intuitive perspective on covalent bonding.

Within quantum alchemy (alchemical perturbation density functional theory), the energy of any molecule \( E_{\text{tot}} \) can be referenced to the uniform electron gas \( E_{\text{UEG}} \) via alchemical thermodynamic integration. Using the Hellmann-Feynman theorem and the chain-rule affords a formally exact atomic energy \( E_I \) partitioning with respect to nuclear charges:

\[
E_{\text{tot}} - E_{\text{UEG}} = \sum_I Z_I \left( \int d\mathbf{r} \frac{\lambda^2}{|\mathbf{r} - \mathbf{R}_J|} \rho(\mathbf{r}, \mathbf{r}) \right) + \frac{1}{2} \sum_I \sum_J \frac{Z_I}{|\mathbf{R}_J - \mathbf{R}_I|},
\]

with \( \rho(\mathbf{r}, \mathbf{r}) \), \( Z_I \), and \( \mathbf{R}_J \) corresponding to the electron density, the nuclear charge, and the position of the nucleus, respectively.

Our DFT based calculated binding energies for varying nuclear charges connecting dozens of chemical bonds among p-block elements reveal simple dependencies of atomic binding contributions on nuclear charges. This has motivated us to introduce the following approximate...
FIG. 1. Experimental measurements (illustrated by atomic force microscopy image of pentacene molecule) inform exact theory (Schrödinger equation & density functional theory) which informs three intuitive approximate views of the chemical bond (MO, VB, QA) in chemical compound space. MO represents molecular orbital theory accounting for bonding in energy diagrams (left). VB corresponds to valence bond theory enabling the decomposition of the wavefunction into covalent and ionic parts (mid). Quantum Alchemy (QA) enables direct partitioning into atomic energies based on thermodynamic integration over varying nuclear charges. MO, VB, and QA can be used to account for bonding trends.

Ansatz for the binding energy \( E_{AB} \) between atoms A and B,

\[
E_{AB} \approx \frac{a}{\text{period’s offset}} - b Z_A Z_B + c \left( \frac{Z_A^{7/3} + Z_B^{7/3}}{\text{free atoms}} \right),
\]

with simple interpretation for each term and requiring just three global parameters \( a, b, c \), which effectively account for interatomic distance and bond order.

After calibration of parameters, we find that this simple model reproduces covalent binding among \( p \)-block elements of either the second, third, or fourth row of the periodic table reasonably well. In the following we also analyse and compare the model to density functional theory, semi-empirical and post-Hartree-Fock quantum mechanics, Pauling’s electronegativity model, and Hammett’s equation. We subsequently discuss limitations, and formulate a generalized expression that predicts bonding trends across multiple periods.

II. RESULTS AND DISCUSSION

A. Performance

After regression of parameters to DFT reference data, our model (Eq. (2)) makes surprisingly accurate estimates. In particular, FIG. 2 shows calculated bond dissociation energies (BDEs) for homolytic cleavage of diatomics A-B saturated with Hydrogens.

\[
H_m A - BH_m \rightarrow H_m A^* + BH_m.
\]

Here, atoms A and B are fourth to seventh main group elements, both of either second, third, or fourth row of the periodic table. BDEs for calibration were obtained from density functional theory (see Computational Details) and the row dependent model parameters were determined from a least-square fit to the ten diatomics in each row (see TABLE I). The residual deviation of estimated binding energies from the DFT reference amounts to an overall mean absolute error (MAE) of just 1.8 kcal/mol across all rows, falling just short of the highly coveted ‘chemical accuracy’ threshold of \( \sim 1 \) kcal/mol. While such accuracy is extremely promising for such a simple functional form and so few parameters, we note the severe restrictions and limitations including fixed effective equilibrium geometries, participating elements, or bond orders.

To set this performance into a wider perspective we compare calculated BDEs among other models well established in the literature. For this purpose, we selected all ten single bond diatomics from the second row contained within the legacy quantum chemistry W4-17 dataset which provides highly accurate bond dissociation energies using explicit electron correlation methods for saturated diatomics in the second row. FIG. 3...
FIG. 2. Calculated bond dissociation energies from density functional theory (DFT, PBE0/def2-TZVP), our quantum alchemy based chemical bond model (Eq. (2)), and Pauling’s electronegativity model in Eq. (4). MAEs with respect to DFT amount to 1.8 and 1.4 kcal/mol for our and Pauling’s model, respectively.

TABLE I. Coefficients and MAEs of our model Eq. (2) for different rows n. b is scaled such that $bZ_A Z_B$ is given in kcal/mol if $Z_A$ and $Z_B$ are given in atomic units.

| n | a (kcal/mol) | b (630/\(a_0\)) | c (kcal/mol) | MAE (kcal/mol) |
|---|---|---|---|---|
| 2 | 215.7 | 10.47 | 1.99 | 1.0 |
| 3 | 392.7 | 8.823 | 1.496 | 2.4 |
| 4 | 1109.2 | 6.180 | 0.813 | 2.2 |

shows a scatter plot of BDEs obtained from various models. Calibrating our model using the W4-17 data yields a leave-one-out prediction error of 1.3 kcal/mol (see Computational Details). Training a chemically agnostic quantum machine learning surrogate model (see computational details) results in a much higher leave-one-out prediction error of 10.4 kcal/mol. By comparison, generic QM methods such as semi-empirical PM7 method, density functional theory (DFT/PBE0/def2-TZVP), and coupled cluster single double perturbative triples F12 calculations, (taken from G2 Ref.32) produce MAEs of 9.5, 1.2, and 0.6 kcal/mol, respectively. This indicates that our model can achieve accurate descriptions of trends in CCS. We should caution, however, that our model is biased due to the calibration, and is likely to perform significantly worse for other chemistries.

B. Comparison to Pauling

Pauling’s model and our model are both simple expression with similar overall accuracy. Pauling’s bond model\(^{10}\) expresses the BDE as

$$E_{AB} = \frac{E_{AA} + E_{BB}}{2} + 23(\chi_A - \chi_B)^2 \quad (4)$$

which depends on the the homolytic binding energies $E_{AA}, E_{BB}$ and the electronegativities $\chi_A, \chi_B$. Pauling’s model is based on a wavefunction decomposition into a covalent and a ionic part. Pauling proposed/assumed/postulated that the bond energy could accordingly be split into a covalent contribution, approximated as the mean of the homolytic bond energies, and an ionic contribution, postulated to be represented by the difference in electronegativities \((\chi_A - \chi_B)^2\). Pauling optimized electronegativities in order to reproduce binding
energies as accurately as possible.\[^{39}\] This model’s predictions for our test sets were discussed before (See also FIGs. 2, 3).

Note that Pauling’s expression requires knowledge of the homolytic bond formation energies and introduces electronegativities as additional quantity while our model relies directly on the nuclear charges. While electronegativity is useful to explain trends in chemical properties, our formulation depending directly on nuclear charges is directly connected to the external potential in the electronic Hamiltonian, and thereby more rigorously rooted in the fundamental physics governing chemistry. Furthermore, homo-diatomics in the third and fourth period dominate the error of our model, while these cases are direct model parameters for Pauling. Note that our model outperforms Pauling’s if the binding partners A and B have a large electronegativity difference as observable for example for the C-F, Si-Cl or N-F bond (FIG. 2).

C. Comparison to Hammett’s equation

Another empirical model to quantify property trends across chemical spaces was proposed by Hammett more than 80 years ago\[^{33,34}\]:

\[
P \approx \rho \cdot \sigma.
\] (5)

Originally, \( P = \log(K_0/K) \) was the equilibrium constant for various reactions of benzene derivatives normalized with respect to a reference reaction constant \( K_0 \). \( \sigma \) described the effect of different substituents and \( \rho \) accounted for the reaction type (e.g. mechanism or solvent). However, the model has been used to describe many other properties like activation energies\[^{35}\], orbital energies of metal organic complexes\[^{36}\] or dipole moment\[^{37}\]. Furthermore, the model has been applied to non-benzyl compounds\[^{38,39}\]. The relation between Hammett constants and electronegativity has also been noted in the context of nucleo- and electro-pilicity relevant for mechanistic discussions in organic chemistry. While Hammett’s model is very intuitive since it only requires separability of two dominating variables, its physical motivation has remained unclear\[^{40}\].

Our model might offer a rationalization of Hammett’s because it emerges from the quantum alchemy based atomic energy decomposition and accounts for the change in binding energy with respect to composition, fully consistent with Hammett’s approach. In particular, the change of binding energy with respect to the nuclear charge, the partial derivative of our model, is simply given by

\[
\frac{\partial E_{AB}}{\partial Z_B}(Z_A, Z_B) = -hZ_A + \frac{7}{3}cZ_B^{4/3}.\] (6)

FIG. 4 displays the near-linear trend of this derivative for a fixed value of \( Z_A \) as a function of the number of valence electrons \( N_{VE} \) of \( Z_B \) for each period investigated in this study.

We have shifted each curve by a constant \( \Delta \) such that it intersects the origin at \( N_{VE} = 0 \). The curves are approximately linear in the number of valence electrons with the slopes being proportional to \( \frac{2}{3}c \) (see Eq. (6)). Thus, \( \frac{\partial E_{AB}}{\partial Z_B} + \Delta \) can be modelled by a Hammett ansatz with \( \sigma = N_{VE} \) and \( \rho \) accounting for different binding behaviour due to a change in number of core electrons for different rows. As the row number increases the slope becomes flatter indicating a lower sensitivity of the binding energy to a change of the binding partner as observable in FIG. 2. The drastic decrease of the slope from the third to the fourth row could be due to the additional 10 3d core electrons for elements in the fourth row.

The identification of the number of valence electrons of binding partner B as the \( \sigma \)-parameter is possible because the influence of the binding partner B is expressed as a function of the nuclear charge in our model. We have arrived at this expression based on the atomic energy decomposition within quantum alchemy. We believe that these findings indicate that such decomposition into atomic or fragment contributions can deepen our understanding of empirical rules such as Hammett’s model. Note that this finding is also consistent with multiple other studies which found a correlation of the \( \sigma \) parameters with atomic quantities like NMR-shifts\[^{41,42}\], polarizing force\[^{43}\], fragment self similarity measures\[^{44}\] or atomic charges\[^{45}\].

D. Limitations to covalent bonds

So far, we have only considered typical covalent bonding scenarios between p-block elements, i.e. elements from groups IV-VII in the periodic table. These elements all have in common that their valence electrons share the same second angular momentum quantum number. Our model has not been developed for decreased cova-
lent character, i.e. for bonding atoms with differing second quantum numbers. Not surprisingly, for example, BDEs for single bonds shown in FIG. 5 indicate a qualitatively different behavior when one bonding partner, say atom B, comes from the alkaline or earth-alkaline group. According to our model calibrated for \( p \) elements only, bond energy content would monotonically increase as the nuclear charge \( Z_B \) decreases. In reality, however, bond energy content must decrease as \( B \) changes from predominantly covalent bonding among \( p \)-block elements into ionic bonding regimes with FF and LiF as the two opposite extremes.

### E. Bond order

FIG. 5 also shows that our model’s results involving Boron (B-N, B-O, and B-F) are worse than for other bonds between \( p \)-block elements. This could be due to the parameters being optimized for bonds with bond orders being close to 1. The B-N, B-O and B-F bonds, however, are known to have a bond order of approximately 1.4. Consequently, our covalent single bond model will systematically underestimate the binding energies for these systems, which is consistent with the numerical observation.

We have investigated whether our model can be recalibrated to also predict binding energies for other bond orders. More specifically, we have refitted parameters \( a, b, c \) to the binding energies of double bonds \( A=B \) for all six possible combinations of carbon, nitrogen and oxygen, saturated with hydrogen. The resulting model also accounts for double bond energies (MAE \( \sim 0.9 \text{ kcal/mol} \)) as long as all reference systems are in the same electronic singlet state. As shown in SI FIG. 10, the inclusion of \( O_2 \) in its triplet ground-state worsens. Using the binding energy of the \( O_2 \) triplet ground state instead, not surprisingly the predictions worsen, in this case such that the MAE becomes 7.4 kcal/mol (SI FIG. 10).

### F. Trends and dependence on period

Since many questions in chemistry only require knowledge about differences in BDEs, we have investigated the applicability of our model towards the prediction of trends among bonds. More specifically, generalization of our model to deal with covalent bonding within \( p \)-block with just 3 parameters \( a, b, c \) is impossible due to the large differences in nuclear charges with increasing row number \( n \). We find, however, that \( a, b, \) and \( c \) vary smoothly with \( n \) when predicting changes in BDEs, i.e. \( \Delta E = E_{AB} - E_{AC} \), in combination with a second order Taylor expansion of the \( Z^{7/3} \)-terms. Then, parameters can be described as simple functions of the principal quantum number \( n \) (the row number) and \( \Delta E \) is approximately given by

\[
\Delta E \approx \Delta Z \left[ 28(n-1) + (8.5 + (n-3)^2)Z_A \right] + 6(Z_B^2 - Z_C^2)
\]

with the change in nuclear charges \( \Delta Z = Z_C - Z_B \) (see Methods section for details).

Encouragingly, Eq. 7 meaningfully reproduces changes in covalent bond energies covering ranges from -40 to +70 kcal/mol with a MAE of just 4.2 kcal/mol with respect to density functional theory. Results are shown in FIG. 6 for any single bond changes within second, third, or fourth row of the periodic table.

Albeit of interest, the generalization of Eq. 7 to also account for atoms A and B coming from different rows, or to involve other bond-orders, has not yet been explored in this study.

---

**FIG. 5.** Limitations of applicability: BDEs between atom A and B corresponding to C/N/O/F or Li/Be/B/C/N/O/F, respectively. For second quantum number differing between A and B, our model [Eq. (2)] breaks down.

**FIG. 6.** Trends in covalent binding: Predicted (Eq. (7)) vs. Truth (corresponding to PBE0/def2-TZVP). 72 binding energy differences between bonds A-B and A-C (\( \Delta E \)), both within either the second, third, or fourth row of the periodic table. \( \Delta E = E_{AB} - E_{AC} \), where \( Z_B \leq Z_C \). Corresponding molecules involved are on display in FIG. 2.
III. CONCLUSION

We have presented a simple and, to the best of our knowledge, novel expression of covalent bonding in terms of nuclear charges. Despite its simplicity, it is deeply rooted in the underlying physics of quantum mechanics via the computational alchemy based reasoning. Our expression might prove useful for developing an improved intuition regarding trends of bonding energies across chemical compound space. It has only three calibration parameters which can easily be regressed to available reference data.

We have found the model to be limited to covalent single and double bonds among atoms with the same first and second quantum number. We have presented promising numerical evidence for p-block atoms (except for rare gas elements) coming from 2nd, 3rd, and 4th row. We have compared our model to Pauling's electronegativity model, and we have discussed its consistency with respect to Hamnett's equation. We note that the description of ionic, metallic, or van der Waals bonding through inclusion of s and d block elements is still outstanding and will be part of future research. This also holds for the generalization to bonds involving elements that differ in principal quantum numbers.

Conceptionally speaking, our model relies on coarse-graining the expectation value of the electronic Hamiltonian throughout chemical space. As such, it is consistent with quantum mechanics and offers a fresh perspective on bond dissociation energies. Historically speaking, it represents an equally powerful yet possibly less empirical alternative to Pauling's electronegativity model.

Future work will show to which extent this partitioning approach can be used to deepen our understanding of chemical space with respect to other extensive properties, and if it is useful for computational materials and molecular design efforts.

IV. METHODS

A. Construction of the model

Continuing from Eq. (1), the electronic energy $E_{\text{el}}^{\text{cl}}$ of a saturated diatomic A-B can be decomposed into

$$E_{\text{el}}^{\text{cl}}_{\text{AB}} = E_{\text{el}}^{\text{cl}}_{\text{A}} + E_{\text{el}}^{\text{cl}}_{\text{B}} + E_{\text{el}}^{\text{cl}}_{\text{AH}} + E_{\text{el}}^{\text{cl}}_{\text{BH}},$$

where $E_{\text{el}}^{\text{cl}}_{\text{A}}$ and $E_{\text{el}}^{\text{cl}}_{\text{B}}$ are the atomic energies of A and B and $E_{\text{el}}^{\text{cl}}_{\text{AH}}$ and $E_{\text{el}}^{\text{cl}}_{\text{BH}}$ is the sum of atomic energies of the hydrogens attached to A and B, respectively (see FIG. 7 exemplary for an example).

Furthermore, the electronic contribution to the binding energy $E_{\text{el}}^{b, \text{cl}}_{\text{AB}}$ is the difference between the energies of the fragments A and B generated by homolytic bond cleavage and the energy of compound AB

$$E_{\text{el}}^{b, \text{cl}}_{\text{AB}} = E_{\text{el}}^{\text{cl}}(\text{A}) + E_{\text{el}}^{\text{cl}}(\text{B}) - E_{\text{el}}^{\text{cl}}(\text{AB}).$$

The atomic binding energy, $E_{\text{el}}^{b, \text{cl}}$, of e.g. atom A is then defined as

$$E_{\text{el}}^{b, \text{cl}}_{\text{A}} = E_{\text{el}}^{\text{cl}}(\text{A}) - E_{\text{el}}^{\text{cl}}(\text{AB})$$

where $E_{\text{el}}^{\text{cl}}(\text{A})$ and $E_{\text{el}}^{\text{cl}}(\text{AB})$ are the atomic energies of A in fragment A and diatomic compound AB, respectively.

Calculating such atomic binding energies according to APDFT, we have observed that $E_{\text{el}}^{b, \text{cl}}_{\text{A}}$ increases approximately linearly with the nuclear charge $Z_{\text{B}}$ of binding partner B for saturated diatomics composed from the elements carbon, nitrogen, oxygen and fluorine (SI FIG. 3).

Furthermore, the atomic energy of the hydrogen atoms remains approximately constant with varying binding partner.

This enables us to express the binding energy approximately as

$$E_{\text{el}}^{b, \text{cl}}_{\text{AB}} \approx \beta_{\text{A}} Z_{\text{B}} + \alpha_{\text{A}} + \beta_{\text{B}} Z_{\text{A}} + \alpha_{\text{B}}.$$ (11)

The energy contributions of fragments A and B are characterized by the parameters $\alpha_{\text{A}}, \beta_{\text{A}}$ and $\alpha_{\text{B}}, \beta_{\text{B}}$, respectively. $\alpha$ accounts for the constant contribution of the heavy atom (A or B) and of the hydrogens attached to it to the binding energy, while $\beta$ describes the contribution of the heavy atom for varying binding partners. Furthermore, we approximate the nuclear repulsion by only considering the interaction between the heavy atoms A and B at an average bond distance $d$. This average is calculated from the bond length between A and B for all considered compounds A-B. Consequently, the total binding energy can be expressed as

$$E_{\text{AB}} \approx \frac{\beta_{\text{A}} Z_{\text{B}} + \alpha_{\text{A}} + \beta_{\text{B}} Z_{\text{A}} + \alpha_{\text{B}} - Z_{\text{A}} Z_{\text{B}}}{d}.$$ (12)

The parameters $\alpha, \beta$ are determined from a least squares fit to $E_{\text{AB}} + \frac{\beta Z}{d}$ for all ten unique combinations A-B of carbon, nitrogen, oxygen and fluorine. The binding energies $E_{\text{AB}}$ are calculated with DFT (see Computational Details).

While Eq. (12) is based on trends for diatomics in the second row of the periodic table, we also apply it to...
diatomics from the fourth to seventh main group of the third and fourth row.

The mean absolute errors (MAEs) for binding energy predictions with Eq. (12) are 0.5, 0.4 and 0.3 kcal/mol for rows 2-4, respectively and the optimized parameters can be found in the SI in TABLE III. The model accurately reproduces the binding energies and can be applied to different rows of the periodic table. However, it is also prone to overfitting because it uses 8 parameters to model binding energies for 10 compounds. Thus, we studied the relation of the parameters α, β on the nuclear charge of the respective element in an attempt to reduce the number of parameters.

The optimized parameter βA correlates linearly with ZA,

\[ \beta_A \approx b' Z_A. \]  

(13)
as shown in SI FIG. 8C. Thus, the terms βAZB, βBZA in Eq. (12) can be written as

\[ \beta_A Z_B + \beta_B Z_A \approx 2b' Z_A Z_B. \]  

(14)
The β-terms account mainly for the large nuclear repulsion term \( Z_A Z_B b^{-1} \gg E_{AB} \) in the dependent variable \( E_{AB} + \frac{Z_A Z_B}{d} \). We note that the linear dependence of \( \beta_A \) on \( Z_A \) is also consistent with the relation

\[ E_{QR}^{el, b} \approx E_{SR}^{el, b} + \frac{1}{2} \left( E_{QQ}^{el, b} + E_{SS}^{el, b} \right), \]  

(15)
between binding energies of elements Q, R and S with \( Z_Q = Z_R - 1 = Z_S - 2 \), that was derived from alchemical enantiomers. Insertion of the definition for electronic binding energies from Eq. (11) into Eq. (15) yields

\[ \beta_{Q} Z_{R} + \alpha_{Q} + \beta_{R} Z_{Q} + \alpha_{R} = \beta_{S} Z_{R} + \alpha_{S} + \beta_{R} Z_{S} + \alpha_{R} \]  
\[ + \beta_{Q} Z_{Q} + \alpha_{Q} - \beta_{S} Z_{S} - \alpha_{S} \]  
\[ \beta_{Q} Z_{R} + \alpha_{Q} + \beta_{R} Z_{Q} + \alpha_{R} = \beta_{S} Z_{R} + \beta_{R} Z_{S} + \beta_{Q} Z_{Q} - \beta_{S} Z_{S} \]  
\[ \beta_{R} (Z_{Q} - Z_{S}) = \beta_{S} (Z_{R} - Z_{S}) + \beta_{Q} (Z_{Q} - Z_{S}) \]  
\[ \beta_{R} = \frac{\beta_{S} + \beta_{Q}}{2}, \]  

(16)
which implies a linear relation between the different values for \( \beta \).

The optimized offset \( \alpha_A \) has a non-linear relationship with \( Z_A \) (SI FIG. 8B). Thus, we model the dependence of \( \alpha_A \) on \( Z_A \) as

\[ \alpha_A \approx c Z_A^{\gamma} + a', \]  

(17)
where the exponent \( \gamma \) accounts for the non-linearity.

By substitution of \( \beta \) and \( \alpha \) in Eq. (12) with Eq. (13) and Eq. (17) and rearrangement of the resulting expression we obtain

\[ E_{AB} \approx a - b Z_A Z_B + c (Z_A^{\gamma} + Z_B^{\gamma}) \]  

(18)
with \( a = 2a' \) and \( b = \left( \frac{1}{2} \right) - 2b' \). The optimization of \( a, b, c, \gamma \) with respect to \( E_{AB} + \frac{Z_A Z_B}{d} \) for each row independently revealed that the optimal exponents \( \gamma_{opt} \) are similar for the different rows (see SI FIG. 9). Thus, \( \gamma \) was kept the same for all rows and only \( a, b, c \) were optimized for each row independently. The optimal value for \( \gamma \) in this optimization procedure was \( \gamma_{opt} = 2.31 \approx \frac{7}{3} \). This is an interesting result because \(-0.768745 \cdot Z_7/3 \) is the leading term in an expansion of the energy of a free atom in its nuclear charge. Hence, the binding energy can be expressed as

\[ E_{AB} \approx a - b Z_A Z_B + c \left( Z_A^{7/3} + Z_B^{7/3} \right) \]  

(19)
with the binding energies of the free atoms \( E_{A}^{atom} \) and \( E_{B}^{atom} \) and \( c' = -0.768745 \).

B. Arriving at Eq. (7)

The \( Z_7/3 \) terms in our model can be approximated by a second order Taylor expansion as

\[ Z_7/3 \approx Z_0^{7/3} + \frac{7}{3} Z_0^{4/3} (Z - Z_0) + \frac{14}{9} Z_0^{1/3} (Z - Z_0)^2 \]  

(20)
with \( Z_0 \) being a suitable reference nuclear charge. Insertion of this expression into our model leads to

\[ E_{AB} = a - b Z_A Z_B \]  
\[ + c \left( \frac{4}{9} Z_0^{7/3} - \frac{7}{9} Z_0^{4/3} (Z_A + Z_B) + \frac{14}{9} Z_0^{1/3} (Z_A^2 + Z_B^2) \right) \]  
\[ = a + \frac{4}{9} Z_0^{7/3} \left( \frac{7}{9} Z_0^{4/3} c (Z_A + Z_B) \right) \]  
\[ - b Z_A Z_B + \frac{14}{9} Z_0^{1/3} c (Z_A^2 + Z_B^2) \]  
\[ = \mu \]  

(21)
The quadratic terms can be written as

\[ Z_A^2 + Z_B^2 = Z_A^3 + Z_B^3 + 2 Z_A Z_B - Z_A Z_B \]  

(22)
leading to

\[ E_{AB} = \xi - \eta (Z_A + Z_B) - b Z_A Z_B \]  
\[ + \mu ((Z_A - Z_B)^2 + 2 Z_A Z_B) \]  
\[ = \xi - \eta (Z_A + Z_B) + (2 \mu - b) Z_A Z_B \]  
\[ + \mu (Z_A - Z_B)^2. \]  

(23)
The energy differences \( \Delta E = E_{AB} - E_{AC} \) between A-B and A-C is then

\[ \Delta E = -\eta (Z_B - Z_C) + \kappa (Z_B - Z_C) \]  
\[ + \mu \left( (Z_A - Z_B)^2 - (Z_A - Z_C)^2 \right) \]  

(24)
TABLE II. Coefficients and MAEs for model Eq. (2) for row \( n = 2, 3, 4 \).

| \( n \) | 0 | \( \eta \) | \( \kappa \) | \( \mu \) |
|---|---|---|---|---|
| 2 | 7.5 | 22.7 | 1.6 | 6.0 |
| 3 | 15.5 | 45.0 | 2.8 | 5.8 |
| 4 | 33.5 | 68.3 | 2.0 | 4.1 |

with \( \kappa = (2\mu - b) \). Values for \( \eta, \kappa, \mu \) are shown in TABLE I for \( Z_0 \) being the mean nuclear charge of the considered elements within one row.

To account for the dependence of \( \eta, \kappa \) and \( \mu \) on the row or principal quantum number \( n \) we make the ansatz \( \eta = \tilde{\eta}(n-1), \kappa = (n-3)^2 - \tilde{\kappa} \), while we choose \( \mu = \tilde{\mu} \) to be independent from the principal quantum number. The parameters \( \tilde{\eta}, \tilde{\kappa} \) and \( \tilde{\mu} \) in the resulting expression

\[
\Delta E = -\tilde{\eta}(n-1)(Z_B - Z_C) \\
+ [\eta(n-3)^2 - \tilde{\kappa}]Z_A(Z_B - Z_C) \\
+ \tilde{\mu}
\]

\[
(Z_A - Z_B)^2 - (Z_A - Z_C)^2
\]

are optimized to produce the minimum MAE for predictions of \( \Delta E \) across rows two to four of the periodic table. The rounded optimized parameters \( \tilde{\eta} = 28, \tilde{\kappa} = 3.5 \) and \( \tilde{\mu} = 6 \) yield the expression

\[
\Delta E = -28(n-1)(Z_B - Z_C) \\
- [8.5 + (n-3)^2]Z_A(Z_B - Z_C) \\
+ 6(Z_A - Z_B)^2 - (Z_A - Z_C)^2
\]

(25)

or alternatively

\[
\Delta E = -28(n-1)\Delta Z - (8.5 + (n-3)^2)Z_A\Delta Z \\
+ 6(Z_B^2 - Z_C^2)
\]

(26)

(27)

with \( \Delta Z = Z_B - Z_C \).

V. COMPUTATIONAL DETAILS

The atomic energies (Eq. 1) of the saturated diatomics were calculated using geometries from the amons dataset\(^{50} \). The fragment structures were generated by splitting the homo-diatomics without further geometry optimization. The required electron densities were obtained following the procedure in earlier work\(^{20} \) from calculations with the CPMD\(^{53} \) code and the CPMD2CUBE program\(^{54} \) using the PBE\(^{55} \) functional in a plane wave basis with a cutoff of 200 Ryd, GTH pseudopotentials and a wavefunction gradient convergence set to \( 10^{-6} \). A primitive cell with a box length of 14.338 \( \text{Å} \) was used for saturated diatomics and of 11.380 \( \text{Å} \) for the fragments. The GTH pseudopotential parameters were scaled by \( \lambda = 6/14, 8/14, 11/14, 1 \) and \( \lambda = 3/7, 4/7, 6/7, 1 \) for the saturated diatomics and the fragments, respectively, to generate electron densities for different values of \( \lambda \). The electron density at \( \lambda = 0 \) was approximated as a uniform electron density. The integration with respect to \( r \) and \( \lambda \) was carried out as summation over grid points and with the trapezoidal rule, respectively.

The binding energies to determine the optimal parameters in Eq. (2) and (12) were calculated with PySCF\(^{59,60} \) with PBE0/def2-TZVP (restricted open shell for fragments). Reported binding energies are for optimized geometries of saturated diatomics and fragments. Initial guesses for the structures were generated through the LERULI API\(^{61} \). The parameter optimization in Eq. (2), (7), (12), (18) was performed by linear or nonlinear least squares fitting as implemented in numpy\(^{57} \) and scipy\(^{58} \). Binding energies of our model presented in FIG. 3 are with respect to W4-17 data after determining the parameters via leave-one-out crossvalidation. The machine learning predictions were obtained from kernel ridge regression with a Gaussian kernel and the bag of bond\(^{59} \) representation in leave-one-out crossvalidation. Predictions with Paulings model Eq. (4) use electronegativities as reported by Pauling\(^{10} \).

Example input files and pseudopotentials to calculate alchemical atomic energies are available at https://github.com/michasahre/bonding_from_alchemy/tree/master. The github repository also contains the alchemical atomic energies, the PBE0 binding energies and a python script to perform the calibration of the parameters in Eq. (2) and (12).

ACKNOWLEDGEMENTS

We acknowledge discussions with M. Meuwly, M. Brah- gato and P. Marquetand, as well as support from the European Research Council (ERCCoG Grant QML). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement #772834.

[1] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Towards the computational design of solid catalysts, Nature Chemistry 1, 37 (2009)
[2] S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, The high-throughput highway to computational materials design, Nature Materials 12, 191 (2013)
SUPPLEMENTARY INFORMATION

FIG. 8. Panel A shows atomic binding energies of atom A as a function of the nuclear charge $Z_B$ of the binding partner for carbon, nitrogen, oxygen and fluorine (solid lines) and the cumulated atomic binding energies of the hydrogens attached to the different atoms A (dashed lines). Panel B shows the optimized parameter $\alpha$ and panel B the optimized $\beta$ as a function of $Z$ after a fit to binding energies from PBE0.
FIG. 9. The mean absolute error for different values of the exponent $\gamma$ in Eq. (18). The curves labeled row 2, row 3 and row 4 show the MAE if the parameters in Eq. (18) are optimized for each row individually. The dotted red curve shows the MAE for all three rows combined if $\gamma$ is kept the same for all rows.

TABLE III. The average bond length $\bar{d}$ of A-B (in Bohr) and the optimized parameters ($\alpha, \beta$) for the elements from main group IV to VII for different rows $n$ of the periodic table. The unit of $\alpha$ is kcal/mol and $\beta$ is given as $630 \cdot ea_0^{-1}$ such that the unit of $\beta A Z A$ is also kcal/mol if $Z A$ is provided in atomic units.

| $n$ | $\bar{d}(a_0)$ | IV     | V       | VI      | VII     |
|-----|---------------|--------|---------|---------|---------|
| 2   | 2.68 (55.45, 703.30) | (92.32, 813.88) | (121.62, 927.30) | (170.25, 1039.70) |
| 3   | 4.01 (-22.56, 1105.16) | (19.86, 1180.19) | (105.45, 1253.62) | (188.06, 1327.44) |
| 4   | 4.46 (-49.06, 2262.88) | (-19.02, 2332.24) | (102.79, 2399.32) | (233.24, 2466.22) |

FIG. 10. Binding energies for double bonds composed from C, N and O. The blue curves show energies if $O_2$ is in the triplet state. The orange curve shows energies if $O_2$ is in the singlet state.