

Quantum Mechanical Treatment of Variable Molecular Composition: From ‘Alchemical’ Changes of State Functions to Rational Compound Design

K. Y. Samuel Chang^a and O. Anatole von Lilienfeld^{*ab}

Abstract: ‘Alchemical’ interpolation paths, *i.e.* coupling systems along fictitious paths without realistic correspondence, are frequently used within materials and molecular modeling simulation protocols for the estimation of changes in state functions such as free energies. We discuss alchemical changes in the context of quantum chemistry, and present illustrative numerical results for the changes of HOMO eigenvalue of the He atom due to alchemical teleportation – the simultaneous annihilation and creation of nuclear charges at different locations. To demonstrate the predictive power of alchemical first order derivatives (Hellmann-Feynman) the covalent bond potential of hydrogen fluoride and hydrogen chloride is investigated, as well as the hydrogen bond in the water–water and water–hydrogen fluoride dimer, respectively. Based on converged electron densities for one configuration, the versatility of alchemical derivatives is exemplified for the screening of entire binding potentials with reasonable accuracy. Finally, we discuss new constraints for the identification of non-linear coupling potentials for which the energy’s Hellmann-Feynman derivative will yield accurate predictions.

Keywords: Alchemical coupling · Computational chemistry · Density functional theory · Potential energy surface · Theoretical chemistry

1. Introduction

Ever since the introduction of Hess’ law and Carnot’s cycle, chemists have known that certain properties, called state functions, always change by the same amount when a system is moved reversibly from one state to another – regardless of how the change has been implemented. The freedom to choose *any* path, even paths without any realistic correspondence except for the endpoints, is exploited within many applications. We generally refer to ‘alchemical’ paths as paths that cannot be followed and verified through experimental observations. For example, Fig. 1(a) illustrates how, according to Hess’ law, the change in enthalpy of reaction can be calculated either by following the (realistic) reaction path, or, just as well, by following a more convenient yet non-realistic (al-

chemical) reaction path to product *via* disassembled elemental states as intermediates. Depending on the choice of state function,

external conditions, system, and process, realistic reaction paths can be very challenging because they can involve many in-

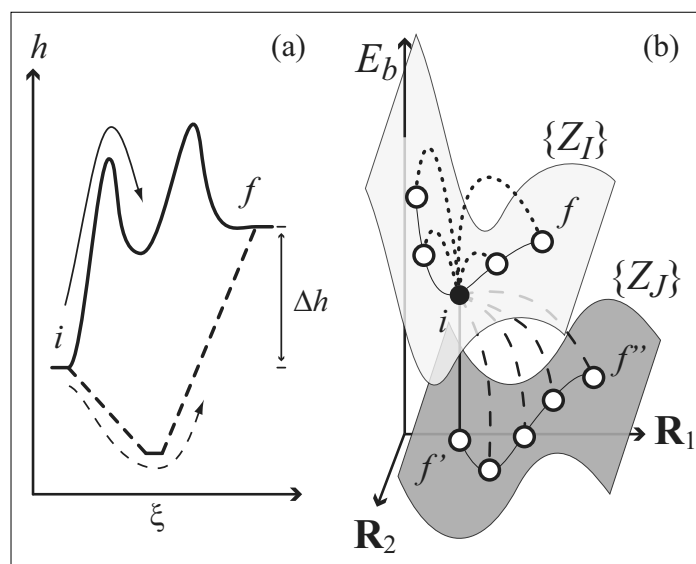


Fig. 1. Alchemical cartoons. (a) The same enthalpy change, Δh , is obtained for a realistic (solid) or an alchemical (dashed) coupling between initial (i) and final (f) states as a function of reaction progress ξ . (b) Alchemical paths connecting compounds on two two-dimensional binding potential energy surfaces corresponding to two different stoichiometries, $\{Z_I\}$ (white) and $\{Z_J\}$ (gray), respectively. Having calculated E_b for some initial system i (filled circle), alchemical paths couple to energies (open circle) of different geometries f with same stoichiometry (dotted), different stoichiometries with same geometry f' (solid), or different geometries and stoichiometries f'' (dashed).

*Correspondence: Prof. O. A. von Lilienfeld
E-mail: anatole.vonlilienfeld@unibas.ch

^aInstitute of Physical Chemistry, NCCR MARVEL
Department of Chemistry
University of Basel

CH-4056 Basel, Switzerland

^bArgonne Leadership Computing Facility,

Argonne National Laboratory, Argonne, IL 0439, USA

intermediate and transition states which are difficult to identify and characterize. Even worse, they might be experimentally impossible to probe, as it is the case for the chemistry of the earth's core, some other planet's bio-sphere, for distant historical or future events, or for very slow or very fast processes.

Within the atomistic theories of quantum and statistical mechanics, any path connecting the Hamiltonian of some initial molecule or material system, H_i , to some final system H_f , can be defined in a coupling order parameter λ as long as the end-points are met,^[1–3] *i.e.*

$$H(\lambda) = \begin{cases} H_i, & \lambda = 0, \\ H_\lambda, & 0 < \lambda < 1, \\ H_f, & \lambda = 1, \end{cases} \quad (1)$$

where $0 \leq \lambda \leq 1$. H_λ in Eqn. (1) denotes some intermediate state at λ , not necessarily differentiable. At boundaries of first order phase transitions, for example, the entropy (state function) is not continuous in temperature (λ). Often, $H(\lambda)$ is (arbitrarily) chosen to be linear in λ , *i.e.* $H(\lambda) = H_i + \lambda(H_f - H_i)$. Thermodynamics textbook examples of such changes include the calculation of the errors made when relying on the ideal gas equation. Introduced as 'computational alchemy'^[2,3] in the realm of computational chemistry, this concept has successfully been used for the interpolation of forces and energies for molecular dynamics (MD) and Monte Carlo (MC) simulations. Also for the purpose of quantum mechanical observables, we can denote any such unrealistic path as 'alchemical'.^[4–6] We note however that Eqn. (1) is also known as 'mutation path' or 'adiabatic connection'.^[7–9]

An even more intriguing possibility for exploiting the freedom of alchemical changes relates to the challenge of rational compound design (RCD). RCD attempts to circumvent (or at least reduce) the combinatorially scaling challenge of having to virtually enumerate and screen larger subsections of chemical or materials compound space using computationally demanding simulation methods. Bottom up approaches to RCD have already been shown to yield promising results for the atomistic control of material, nanoparticle, and even molecular structures.^[10,11] Because of the vastness of chemical compound space (CCS), the virtual identification of novel compounds that meet desired property requirements still remains a challenge.^[5,12] Once an alchemical interpolating path, $H(\lambda)$, is defined, property derivatives with respect to λ can be evaluated^[13] (see Section 4). Similar to an iterative gradient descent-like algorithm, one could

thus navigate gigantic combinatorial compound libraries at dramatically reduced computational costs by only visiting the most promising compounds, one after the other, while avoiding the least promising candidates.^[14,15]

The concept of connecting different systems *via* Eqn. (1) has been in frequent use in various research fields, including computational engineering, physics, biophysics, and chemistry. Here, we first briefly summarize the most common application of Eqn. (1) to calculate free energy changes, or alloy formation energies in Section 2. In Section 3 we review the quantum mechanical treatment of alchemical changes. To this end, we mainly rely on the use of density functional theory (DFT) even though analogous arguments can be made using conventional wave-function based quantum chemistry methods. In Section 4 we present numerical results that demonstrate the use of alchemical derivatives for screening entire potential energy binding surfaces with semi-quantitative accuracy and without additional self-consistent field calculations.

2. Common Alchemical Applications

Free energy is one of the most important state functions in chemistry. Since it is a statistical average, large numbers of configurations need to be taken into account to yield accurate predictions.^[16] For example, calculating a free energy of solvation following a path that mimics the realistic complex process of reversible microscopic immersion of the solute into a condensed ensemble of a very large number of solvent molecules would imply a severe simulation effort to ensure that all relevant degrees of freedom have sufficiently been sampled. Furthermore, to account for hysteresis effects, this simulation should be repeated for various initial conditions and immersion rates. And one would have to start anew for any changes made to temperature, pressure, or solvent and solute species. Alternatively, one could also calculate the *change* in free energy with respect to some solute for which the free energy of solvation is already known. Thermodynamic integration, *i.e.* numerical integration of the statistical mechanical average of the 'alchemical force' along the path converting known solute ($\lambda = 0$) into query solute ($\lambda = 1$),^[3] yields

$$\Delta G = \int_0^1 d\lambda \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda. \quad (2)$$

Jorgensen and Ravimohan^[7] proposed an even more efficient alternative: One can

also estimate the change in free energy of solvation due to changing the solute using perturbation theory and MC simulation. Specifically, they considered the effect on the free energy of hydration due to an alchemical change of a methyl into hydroxy-group, $\Delta G = G_f - G_i = G_{\text{CH}_3\text{CH}_3} - G_{\text{CH}_3\text{OH}}$. One can show that if the sampling of the two states, H_i and H_f , yields sufficient overlap, the corresponding free energy difference can be accurately predicted using perturbation theory,

$$e^{-\beta \Delta G} \approx \left\langle e^{\beta(H_f - H_i)} \right\rangle_i. \quad (3)$$

Here, $1/\beta = k_B T$, and the right-hand-side refers to the average of the Hamiltonian difference Boltzmann's weight over a trajectory generated using H_i . The authors used a linear interpolation of force field parameters for methanol and ethane, $H(\lambda) = H_{\text{CH}_3\text{OH}} - \lambda(H_{\text{CH}_3\text{CH}_3} - H_{\text{CH}_3\text{OH}})$, from which the energy can be calculated for any λ .

As such, alchemical changes enable the prediction of changes in free energy differences without having to actually model the realistic process under investigation. Linear interpolation approaches have been applied to free energy calculations in various chemical and biological systems.^[17–20] Smith and van Gunsteren found that non-linear alchemical coupling does not necessarily lead to linear free energy changes.^[21] Further applications of alchemical coupling to the estimation of free energy difference include the free energy of hydration of ions using *ab initio* molecular dynamics,^[22] differences in free energy of binding between various host-guest complexes,^[23] free energy differences at phase boundaries to predict melting points,^[24,25] the free energy of mixing to identify eutectics in ternary mixtures of molten alkali-nitrate salts,^[26] kinetic isotope effects,^[27] as well as constraints on the composition of the Earth's core.^[28]

But also from the solid state point of view the concept of alchemical coupling can be used for the prediction of properties of disordered materials, such as co-crystals, solid solutions, or solid mixtures, as a function of mole-fraction.^[29] It is computationally difficult to deal with such mixed disordered systems since the minimal self-repeating units can become very large. As a result it is nearly impossible to set up disordered systems within periodic boundary conditions. One alternative consists of using cluster-expansion methods.^[30] Another alternative, akin to alchemical coupling, is the virtual crystal approximation (VCA)^[31] which averages the system, rather than explicitly representing it. One of the simplest disordered class of materials are ter-

nary semiconductors, $A_xB_{1-x}C$, where AC and BC are two different semiconductors while x is the mole-fraction between A and C. Consider, for example,^[32] Eqn. (1) applied to AlGaAs; $Al_xGa_{1-x}As$: $H(x) = H_{GaAs} + x(H_{AlAs} - H_{GaAs})$. The linear interpolated alchemical path describes an averaged Hamiltonian between AIAs and GaAs for any mole-fraction of Al and Ga.

3. Alchemy in Quantum Mechanics

3.1 Fictitious Systems

Within a first principles notion of CCS,^[12] one can view every compound in any geometry as a micro-state described by a unique electronic Hamiltonian H . More specifically, the total potential energy's molecular Hamiltonian, H , is a function of a given set of nuclear coordinates, charges, and number of electrons, $\{\mathbf{R}_p, Z_p, N_e\}$, respectively. Without any loss of generality, we here rely on the Born-Oppenheimer approximation, neglecting all non-adiabatic electronic, relativistic or nuclear quantum effects. Studies of alchemical paths have historically provided essential insight into the density functional theory (DFT) formulation of the many-electron problem in molecules.^[33,34] In 1974, Harris and Jones introduced an adiabatic connection,^[35] coupling the system of interest to a fictitious yet relevant system of non-interacting electrons,

$$H(\lambda) = T + \lambda V_{ee} + V_{ext}, \quad (4)$$

where T , V_{ee} , and V_{ext} represent the kinetic energy, electron–electron interaction energy, and external potential energy operators. By changing λ , one can dial in the electron–electron interaction. For $\lambda = 0$, the electronic Schrödinger equation can be solved analytically, providing useful information on properties such as the exchange–correlation hole,^[36–38] an important ingredient for current exchange–correlation potential development efforts.^[39–42] Another important study of electron–electron interaction, carried out by Seidl, Perdew and Levy, introduces the limit of strictly correlated electrons.^[43]

E. B. Wilson introduced the idea to alchemically couple any system to the uniform electron gas. Based on this path, he derived an expression for an exact four-dimensional density functional theory, integrating over three spatial and one λ -dimension.^[37,44] Subsequently, Politzer and Parr^[45] showed that, by defining free-atom screening functions, Wilson's functional can be decomposed into kinetic and potential energy of N_e electrons. These definitions of DFT related alchemical

paths constitute the underlying framework for the results and discussions presented in what follows.

Within ground-state DFT,^[33] we can easily calculate $E(\lambda)$ for any iso-electronic change of geometry and composition, *i.e.* under the constraint that $\int d\mathbf{r} n_\lambda(\mathbf{r}) = N_e \forall 0 \leq \lambda \leq 1$,

$$E[n_\lambda] = T[n_\lambda] + V_{ee}[n_\lambda] + \int d\mathbf{r} n_\lambda(\mathbf{r}) v_{ext}(\mathbf{r}, \lambda). \quad (5)$$

Here, the coupling is introduced explicitly and exclusively through the external potential. In practice, such coupling can be realized by scaling up or down the pseudopotentials or nuclear charges of initial and final molecules at their distinct clamped geometries. Note that kinetic and potential energy terms are only implicitly dependent on λ , namely through the electron density's dependency on the λ -dependent external potential – which is imposed through application of the variational principle.

3.2 Alchemical Teleportation of an Atom

To illustrate the idea of alchemical changes within quantum chemistry, we now consider a process which is trivial when done through a realistic path, and non-trivial when done alchemically: The ‘teleportation’ of an atom from one site to another with the constraint that the total number of electrons and protons is kept constant. Thus, instead of the trivial real space displacement of the atom, we continuously decrease the nuclear charge (annihilation) at one site while continuously increasing (creation) the nuclear charge at the other site by the same amount. For example, the external potentials of an atom at two sites can be linearly coupled through an alchemical path,

$$H(\lambda) = T + V_{ee} + Z \sum_i^{N_e} \left(\frac{1-\lambda}{|\mathbf{r}_i|} + \frac{\lambda}{|\mathbf{r}_i - \mathbf{R}|} \right), \quad (6)$$

where the respective atomic sites are located at the origin and at \mathbf{R} . Considering only the endpoints ($\lambda = (0,1)$), the location of the atom obviously shifted from origin to \mathbf{R} . For any intermediate value of λ , however, the electrons will distribute among the two competing poles of the external potential given in Eqn. (6), forming an attractive chemical bond.

To numerically exemplify this process, we have chosen the highest occupied molecular orbital (HOMO) eigenvalue, ϵ , as property of interest, and an alchemical change corresponding to the linear teleportation of a $Z = 2$ and $N_e = 2$ system, *i.e.* effectively translating the He atom.

The numerical calculation of ϵ for variable λ has been carried out using interpolated pseudopotentials within plane-wave basis set PBE DFT calculations, in analogy to previous studies.^[5,13,14] See Section 7 for more details.

In Fig. 2(a), the λ -dependency of ϵ is shown for various distances between the two atomic sites, $d = |\mathbf{R}|$. Clearly, while alchemical paths for small d yield simple parabolic shapes of ϵ , for teleportation involving larger interatomic distances ϵ develops into a double hill. ϵ versus d is plotted in Fig. 2(b) for various λ values. We note that for $\lambda = 0.5$ (magenta), the d dependency of ϵ corresponds to the case of stretching H_2 . ϵ increases monotonically at $\lambda = 0.1$ and $\lambda = 0.2$ as d increases. For these λ values, the buildup of integrated electron density at the \mathbf{R} , is still negligible, Fig. 2(c). Overall, the effect of nuclear potential in Eqn. (6),

$$\frac{2\lambda}{|\mathbf{r}_i - \mathbf{R}|},$$

amounts to a static electric field, which appears to induce a Stark effect.^[46–48] Since the electric field decreases according to Coulomb's law, ϵ rises as a result of decreasing electric field perturbation. Apart from the delocalization error of DFT,^[49,50] such nonlinear behavior could also be related to the instability of H_2^+ -like systems, which has been shown analytically.^[51] Hogreve pointed out that the strongly polarized electron density of an asymmetric H_2^+ -like molecule severely destabilizes the system.^[52] While the additional electron stabilizes our system, nonlinear behavior can be expected for ϵ in the case of strongly polarized density, *i.e.* for $\lambda > 0.3$. Fig. 2(c) displays integrated electron density slices, $\Delta(z) = \int dx dy n(x,y,z)$, for various λ values at interatomic distance, $d = 5 \text{ \AA}$. Note that for $\lambda = 0.5$, the electron density distribution corresponds to H_2 . The non-linear dependency of electron density n on linearly changing growth of nuclear charge can be seen in Fig. 2(d) for the abrupt changes in electron density response induced by going from $\lambda \approx 0.2$ to $\lambda \approx 0.3$. To investigate the impact of parameterized exchange correlation potentials in DFT, Cohen and Mori-Sánchez calculated similar changes for $N_e = 1$ and $N_e = 2$ using the hydrogen atom plus one additional atomic site where a nuclear charge is grown, *i.e.* $Z(\lambda)$ with $Z(\lambda = 0) = 0$, $Z(\lambda = 0.5) = 1$ (H), and $Z(\lambda = 1) = 2$ (He).^[53]

4. Rational Compound Design

4.1 Motivation

The goal of RCD corresponds to solving the inverse question, *i.e.* “which compounds exhibit a set of pre-defined desired properties?”, at a rate that is superior to mere screening.^[5,54–57] Various

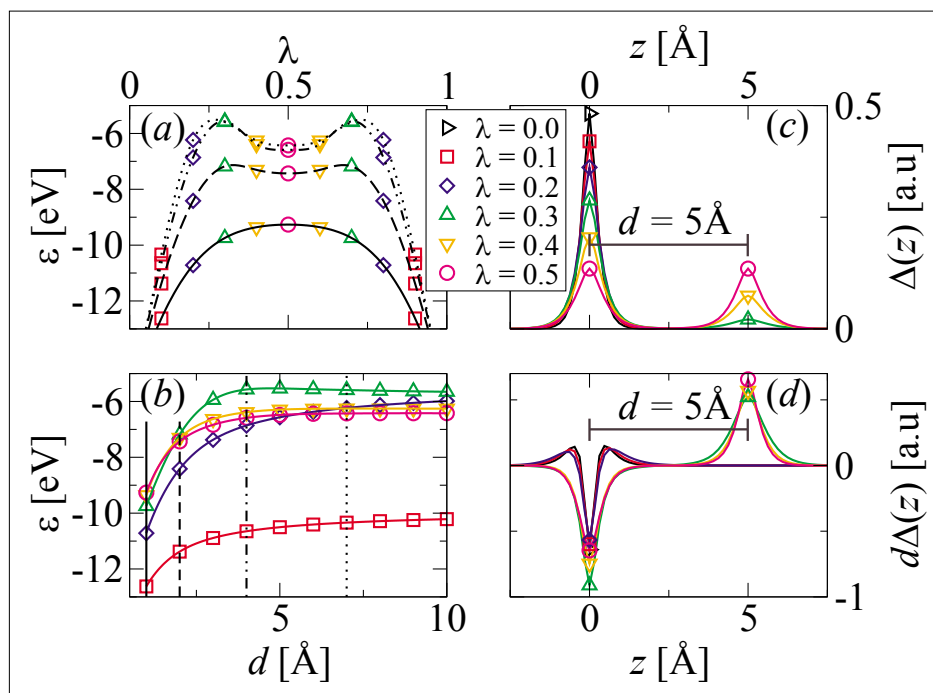


Fig. 2. Alchemical transportation of a He atom. (a) ϵ as a function of λ for four distances $d \in \{1, 2, 4, 7\}$ Å denoted by solid, dashed, dash-dotted, dotted lines, respectively. (b) ϵ as a function of d . (c) Integrated electron density, $\Delta(z) = \int dx dy n_\lambda(x, y, z)$ for various λ at $d = 5$ Å aligned with z . The electronic cusps at the nuclear sites have been highlighted by their corresponding λ symbols. (d) Integrated response of electron density due to changing λ , $d\Delta(z) = \int dx dy \partial_\lambda n_\lambda(x, y, z)$ for various λ at $d = 5$ Å aligned with z .

approaches tackle this problem, including the inverse spectrum approach,^[58] linear combination of atomic potentials,^[59,60] and many others.^[56,61–63] For the electronic potential energy, an alchemical path coupling $E(\lambda = 0)$ of one molecule to an unknown $E(\lambda = 1)$ of another compound makes explicit the compositional dependence of the energy. Understanding such a dependence holds promise to dramatically reduce the computational burden of having to stubbornly screen one compound after the other. More specifically, we can expand E in λ in terms of a Taylor series,

$$E(\lambda) = E_i + \lambda \partial_\lambda E_i + \frac{\lambda^2}{2} \partial_\lambda^2 E_i + \dots, \quad (7)$$

where the subscript of E_i represents the quantum mechanical expectation value of H_i . In other words, if all derivatives of E_i were available one could simply follow a steepest descent procedure to screen a set of coupled ‘neighboring’ molecules, *e.g.* with small differences in geometry or stoichiometry, to identify and proceed to more promising compound candidates. Fig. 1(b) illustrates the exploration of CCS following such alchemical predictions. Ideally, only a single calculation of the electronic ground-state E_i would be required (denoted by black circle). The energy of neighboring compounds (denoted by white circles)

can then be estimated *via* Eqn. (8). As we discuss below, it is possible to make such scans through changes in geometry as well as composition.

In Ref. [13] we already discussed that for *any* iso-electronic alchemical change, the first order derivative is simply the Hellmann-Feynman derivative.^[64] Consequently, differentiation of Eqn. (5) yields,

$$\begin{aligned} \partial_\lambda E[n_\lambda, \lambda] &= \langle \partial_\lambda H \rangle_\lambda \\ &= \int dr n_\lambda(\mathbf{r}) \partial_\lambda v_{ext}(\mathbf{r}, \lambda), \quad (8) \end{aligned}$$

which is the same as the first order perturbation term.^[46] Higher order derivatives can be evaluated or approximated by linear response theory,^[65–67] and will be discussed below in the context of linearizing the energy in λ in Section 4.4.

4.2 Alchemical Changes in Geometry

We now consider alchemical changes that only involve teleportation. To demonstrate the versatility and transferability of the discussed approach, we have calculated alchemical predictions of changes in binding energy for two very different modes of binding: The covalent interatomic potential in hydrogen fluoride, as well as the hydrogen-bond-dominated van der Waals potential of the water dimer. In both cases the binding energy is

given as the difference in potential energy of dimer (dim) and (relaxed) monomers $m1$ and $m2$, $E_b(d) = E_{\text{dim}}(d) - E_{m1} - E_{m2}$. Any approximate solution of the electronic Schrödinger equation at some initial distance d_i enables us to estimate the binding energy of any other d using the Hellmann-Feynman derivative and first order Taylor expansion in the alchemical teleportation path (Eqns. (8,7)),

$$E_b(d) \approx E_b^{T1}(d) = E_b(d_i) + \partial_\lambda E_b(d_i). \quad (9)$$

Considering now the case of d_i corresponding to the equilibrium distance, d_{eq} , the insets of the two top panels in Fig. 3 show the resulting scatter plots of $E_b^{T1}(d)$ versus the actual $E_b(d)$ for various values of d in the case of HF and $(\text{H}_2\text{O})_2$. While there is clear correlation, the scale differs dramatically for the two modes of binding. Most importantly, in the case of the dissociative tail E^{T1} correlates practically linearly with the actual binding energy. Consequently, if we now approximate the true $E_b \approx E_b^p = a_{lr} E_b^{T1} + b_{lr}$, (l and r correspond to the left-hand repulsive wall and the right-hand attractive tail, respectively) one can solve for the coefficients a and b if further constraints are known. Since this is a rather exploratory study, we here simply assume that (i) $E_b(d = d_{eq}) = E_b^{T1}(d_{eq})$, and (ii) $E_b(d \rightarrow \infty) = 0$ in the dissociative region of the curve, and (iii) in the case of the repulsive region that $E_b(d = \frac{2}{3} d_{eq}) = 0$ for covalent binding, and $E_b(d = \frac{5}{6} d_{eq}) = 0$ for intermolecular binding. Assumption (iii) is based on experience using typical Morse and Lennard-Jones parameters, respectively. All resulting coefficients $\{a_{lr}, b_{lr}\}$ are specified in Ref. [68]. The predictions for scanning the entire binding potential agree reasonably well with the true binding potentials, and are shown together for both systems in the top panels in Fig. 3. Integrated deviations of these predictions are also shown in Table 1, yielding single digit percentage error for predicting the integral over the covalent bonding potential of hydrogen fluoride, and $\sim 14\%$ error for the integral over the van der Waals potential of the water dimer. We stress that the *entire* screen using this model only requires a single self-consistent field cycle to calculate energy and derivatives at $d = d_{eq}$.

4.3 Alchemical Changes in Stoichiometry

We now extend the use of Eqn. (9) to also make predictions not only for teleportation changes in geometry but also for transmutational changes in stoichiometry. In particular, we have calculated predictions for changing hydrogen fluoride into

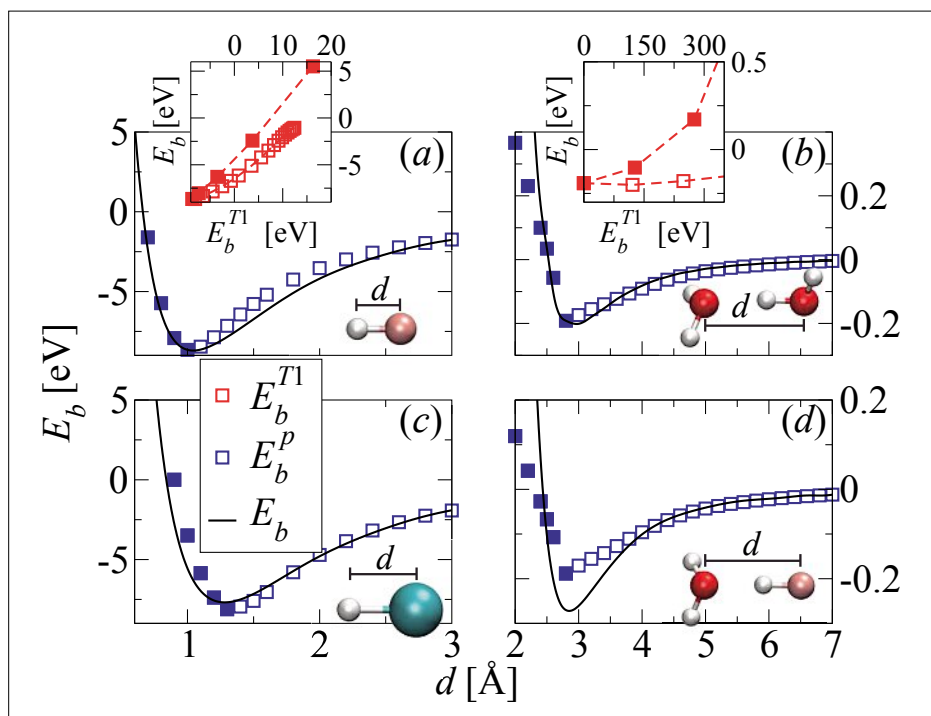


Fig. 3. Actual (black lines) and alchemical (blue squares) binding energy E_b of repulsive (filled) and attractive (empty) regions of binding potentials for HF (a), $(\text{H}_2\text{O})_2$ (b), HCl (c), and H_2O -HF (d). Each screen corresponds to using only one self-consistent field (SCF) calculation at $d_i = d_{eq}$, together with the first order Taylor-expansion based model, $E_b^p = a_{lr} E_b^{T1} + b_{lr}$ (Eqn. (7)). Insets in (a) and (b) show E_b versus E_b^{T1} . The screens in (c) and (d) are slightly less predictive because they are made using SCF results from HF and $(\text{H}_2\text{O})_2$, respectively. d_{eq} is set to 1, 2.8, 1.4 and 2.8 Å for (a), (b), (c), and (d) respectively.

Table 1. Numerical integrals of reference energies E_b (REF) and of absolute deviation of alchemical predictions E_b^p from reference energies E_b (PRE-REF) over the entire binding region, *i.e.* for all d where $E_b < 0$, and percentage thereof (%) for the repulsive wall predictions as well as for the attractive tail. Columns correspond to (a) HF, (b) $(\text{H}_2\text{O})_2$, (c) HCl, and (d) H_2O -HF on display in Fig. 3.

Integral [eV×Å]	(a)	(b)	(c)	(d)
REF[eV×Å] (wall)	-2.542	-0.059	-3.054	-0.095
PRE-REF[eV×Å] (wall)	0.150	0.009	0.656	0.036
% (wall)	5.9	15.6	21.5	37.6
REF [eV×Å] (tail)	-8.199	-0.275	-6.017	-0.594
PRE-REF [eV×Å] (tail)	0.692	0.031	0.239	0.072
% (tail)	8.4	11.4	4.0	12.1

hydrogen chloride at various interatomic distances, as well as changing the water dimer into the water–hydrogen fluoride complex. Since we use pseudopotentials for both of these changes the total number of valence-electrons in our calculations does not change. To calculate E_b^{T1} according to Eqn. (9) we have chosen d_{eq} to correspond to the equilibrium distance of the target system, *i.e.* HCl and H_2O -HF. Again, the same assumptions (i)–(iii) as above are used to calculate a_{lr} and b_{lr} to obtain a linear approximation of the actual $E_b(d)$ in E_b^{T1} . Also for these changes, the resulting coefficients are specified in Ref. [68]. The predicted binding curves show reasonable agreement with the actual numbers, as shown for both systems in the bottom

panels in Fig. 3. Again, integrated and relative errors are given in Table 1, and show a reasonable albeit slightly worse performance than in the case of predicting the water dimer or the hydrogen fluoride. We reiterate, however, that the entire screen results from only one self-consistent field cycle carried out to calculate energy and derivative of *another* molecular system – at the d_{eq} of the target system. While it is also possible to use other d to calculate energies and derivatives this typically leads to less accurate predictions. We do not think that this constitutes a problem since knowledge about equilibrium distances of target structures can easily be obtained from literature or through inexpensive force-field or semi-empirical quantum chemistry calculations

which incur negligible computational overhead.

4.4 Linearizing Chemical Space

As we have seen above for the teleportation of the He atom, as well as in other studies,^[12,13] there are cases when the first order Taylor expansion of Eqn. (9) does not provide satisfactory predictive power. This is not surprising since changes in composition correspond to large perturbations that typically lead to non-linear responses. We believe that the good performance obtained above for the binding curves is due to cancellation of higher order effects and due to the calibration of the linear model to the appropriate physical dissociation or repulsion limits. One way to systematically improve the predictive accuracy consists of including increasingly higher-order terms. Sebastiani and coworkers^[67,69,70] as well as Geerlings, De Proft and others^[71–73] proposed promising approaches in this direction. For example, akin to our discussion above, Benoit, Sebastiani and Parrinello investigated the performance of second order linear response theory for screening the potential energy surface of the water dimer, and achieved very high predictive power.^[74] How to efficiently calculate susceptibility accurately and in general, however, is still a matter of current research. Furthermore, typically one observes a (sometimes dramatic) increase in computational cost due to λ -dependent susceptibilities, thereby defying the original motivation of RCD to navigate CCS *without* having to solve Schrödinger's equation from scratch for each and every new geometry or molecule. As pointed out in Ref. [13], a promising alternative route towards improving the predictive power of the first order derivative consists of deviating from the assumption that the alchemical coupling must be linear in λ . In fact, as already mentioned above in the context of interpolating force-fields,^[21] we are free to use *any* kind of coupling as long as we meet our endpoints, *i.e.* comply with Eqn. (1). More specifically, if we knew the form of some coupling external potential $v_{ext}(\mathbf{r}, \lambda)$ that induces such changes in the electron density that $E(\lambda)$ becomes linear in λ , then Eqn. (9) would result in perfect predictions. The quest for such a potential has been discussed in Ref. [12], in particular in connection to a 1-ounce-of-gold prize for anyone who provides a practical solution to this problem.

For a coupling path to generally fulfill the requirement that $E(\lambda)$ becomes linear in λ we note that the potential must have such a shape that the first order derivative, $\partial_\lambda E$ is a constant (as already pointed out and used in ref. [13]), and that furthermore, all higher order energy derivatives must be zero. Consequently,

$$0 = \partial_{\lambda}^m E[n_{\lambda}] = \int dr \partial_{\lambda}^{m-1} (n_{\lambda}(\mathbf{r}) \partial_{\lambda} v_{ext}(\mathbf{r}, \lambda)), \quad (10)$$

$\forall m > 1$. This imposes certain constraints on the interpolating potential. For example, in the case of the second order derivative, equating the integrand to zero and solving for the electron density's response results in

$$\partial_{\lambda} n(\mathbf{r}) = -n_{\lambda}(\mathbf{r}) \frac{\partial_{\lambda}^2 v_{ext}(\mathbf{r}, \lambda)}{\partial_{\lambda} v_{ext}(\mathbf{r}, \lambda)}. \quad (11)$$

Similar expressions can be obtained for higher order density response functions. Possibly, Eqn. (10) could be transformed into a variational problem that yields an interpolating potential with the desired effect that the associated energies are indeed linear in λ .

5. Conclusions

We discussed recent theoretical developments and approaches based on coupling states using unrealistic 'alchemical' paths. Numerical evidence has been presented for the applicability and versatility of alchemical approaches applied to the inexpensive prediction of quantum mechanical observables of novel systems. The derivative based predictions certainly reflect the qualitative trend of the desired binding potentials, and are within single, or low double, digit percentage accuracy. Results, discussions, and current state of the field indicate that the study of generalized coupling approaches still holds great promise for the predictive simulation of molecular and materials properties, as well as for rational compound design.

6. Computational Details

All calculations have been carried out using Kohn-Sham DFT^[75] as implemented in CPMD^[76] with PBE (He teletransportation) or PBE0 (all other calculations) functional.^[40,77] Goedecker pseudopotentials^[78–80] have been used as published by Krack,^[81] in conjunction with 100 Ry plane-wave cutoffs in isolated $30 \times 15 \times 15 \text{ \AA}^3$ box for He, $20 \times 20 \times 20 \text{ \AA}^3$ for HF \rightarrow HF and HF \rightarrow HCl, and 110 Ry plane-wave cutoff with isolated $25 \times 15 \times 15 \text{ \AA}^3$ box for H₂O \rightarrow H₂O and H₂O \rightarrow HF. Alchemical coupling has been imposed through linear interpolation of corresponding pseudopotential parameters,^[5,13,14] $\sigma(\lambda) = \sigma_i + \lambda(\sigma_f - \sigma_i)$, where σ_i and σ_f represent the parameters for atoms in H_i and H_f respectively. HOMO eigenvalues ε have been calculated

as a finite difference relying on Jana'k and Koopman's theorem,^[82,83] $\varepsilon \approx \frac{E_{N+\delta} - E_N}{\delta}$ where δ is 1% of a positive unit charge. The geometry scans of HF and HCl have been performed by fixing the heavy atoms at origin, moving H in d direction, while aligning the HF or HCl bond along z -axis. In the case of the (H₂O)₂ scan, equilibrium geometry has been taken from S66 data base,^[84] setting the oxygen of the H-acceptor at the origin, while aligning the O-H bond of the H-donor with the z -axis. The H₂O-HF geometry scans have been performed by replacing the oxygen of the H-donor by F and annihilating the other hydrogen while keeping the HF bond aligned with the z -axis.

Acknowledgments

Both authors acknowledge funding from the Swiss National Science foundation (No. PPOOP2_138932). This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. DOE under contract DE-AC02-06CH11357.

Received: May 19, 2014

- [1] W. G. van Gunsteren, H. J. C. Berendsen, *J. Comput. Aided Mol. Des.* **1987**, *1*, 171.
- [2] T. Straatsma, J. A. McCammon, *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.
- [3] P. Kollman, *Chem. Rev.* **1993**, *93*, 2395.
- [4] N. Marzari, S. de Gironcoli, S. Baroni, *Phys. Rev. Lett.* **1994**, *72*, 4001.
- [5] O. A. von Lilienfeld, R. D. Lins, U. Rothlisberger, *Phys. Rev. Lett.* **2005**, *95*, 153002.
- [6] O. A. von Lilienfeld, M. E. Tuckerman, *J. Chem. Phys.* **2006**, *125*, 154104.
- [7] W. L. Jorgensen, C. Ravimohan, *J. Chem. Phys.* **1985**, *83*, 3050.
- [8] D. C. Langreth, J. P. Perdew, *Phys. Rev. B* **1980**, *21*, 5469.
- [9] J. Harris, *Phys. Rev. A* **1984**, *29*, 1648.
- [10] J. V. Barth, G. Costantini, K. Kern, *Nature*, **2005**, *437*, 671.
- [11] Y. Xia, Y. Xiong, B. Lim, S. E. Skrablak, *Angew. Chem. Int. Ed.* **2009**, *48*, 60.
- [12] O. A. von Lilienfeld, *Int. J. Quantum Chem.* **2013**, *113*, 1676.
- [13] O. A. von Lilienfeld, *J. Chem. Phys.* **2009**, *131*, 164102.
- [14] D. Sheppard, G. Henkelman, O. A. von Lilienfeld, *J. Chem. Phys.* **2010**, *133*, 084104.
- [15] S. Goedecker, *J. Phys. Chem.* **2004**, *120*, 9911.
- [16] M. E. Tuckerman, 'Statistical mechanics: Theory and molecular simulation', Oxford University Press, **2010**.
- [17] W. L. Jorgensen, *Science* **2004**, *303*, 1813.
- [18] C. Oostenbrink, W. F. van Gunsteren, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6750.
- [19] S. Riniker, C. D. Christ, H. S. Hansen, P. H. Hüsenberger, C. Oostenbrink, D. Steiner, W. F. van Gunsteren, *J. Phys. Chem. B* **2011**, *115*, 13570.
- [20] D. Jiao, K. Leung, S. B. Rempe, T. M. Nenoff, *J. Chem. Theory Comput.* **2011**, *7*, 485.
- [21] P. E. Smith, W. G. van Gunsteren, *J. Phys. Chem.* **1994**, *98*, 13735.
- [22] K. Leung, S. B. Rempe, O. A. von Lilienfeld, *J. Chem. Phys.* **2009**, *130*, 204507.
- [23] C. Oostenbrink, *J. Comp. Chem.* **2009**, *30*, 212.
- [24] S. Jayaraman, E. J. Maginn, *J. Chem. Phys.* **2007**, *127*, 214504.
- [25] S. Jayaraman, A. P. Thompson, O. A. von Lilienfeld, E. J. Maginn, *Ind. Eng. Chem. Res.* **2010**, *49*, 559.
- [26] S. Jayaraman, A. P. Thompson, O. A. von Lilienfeld, *Phys. Rev. E* **2011**, *84*, 030201.
- [27] A. Pérez, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **2011**, *7*, 2358.
- [28] D. Alfé, M. J. Gillan, G. D. Price, *Nature* **2000**, *405*, 172.
- [29] S. Adachi, *J. Appl. Phys.* **1985**, *58*, R1.
- [30] T. Mueller, G. Ceder, *Phys. Rev. B* **2009**, *80*, 024103.
- [31] L. Bellaiche, D. Vanderbilt, *Phys. Rev. B* **2000**, *61*, 7877.
- [32] D. G. Pettifor, *Acta Mater.* **2003**, *51*, 5649.
- [33] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.
- [34] R. G. Parr, W. Yang, 'Density functional theory of atoms and molecules', Oxford Science Publications, **1989**.
- [35] J. Harris, R. O. Jones, *J. Phys. F: Metal Phys.* **1974**, *4*, 1170.
- [36] O. Gunnarson, B. I. Lundqvist, *Phys. Rev. B* **1976**, *13*, 4274.
- [37] R. G. Parr, W. Yang, 'Density-Functional Theory of Atoms and Molecules', Oxford University Press, USA, **1994**.
- [38] W. Koch, M. C. Holthausen, 'A Chemist's Guide to Density Functional Theory', WILEY-VCH, Germany, 2nd ed., **2002**.
- [39] J. P. Perdew, A. Zunger, *Phys. Rev. B* **1981**, *23*, 5048.
- [40] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
- [41] M. Levy, *Phys. Rev. A* **1991**, *43*, 4637.
- [42] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais, *Phys. Rev. B* **1992**, *46*, 6671.
- [43] M. Seidl, J. P. Perdew, M. Levy, *Phys. Rev. A* **1999**, *59*, 51.
- [44] J. E. Bright Wilson, *J. Phys. Chem.* **1962**, *36*, 2232.
- [45] P. Politzer, R. G. Parr, *J. Phys. Chem.* **1974**, *61*, 4258.
- [46] D. Griffiths, 'Introduction to Quantum Mechanics', 2nd ed., **2005**.
- [47] M. Okrus, R. Müller, A. Hese, *J. Chem. Phys.* **1999**, *110*, 10393.
- [48] W. Xue, M. A. Ratner, *Phys. Rev. B* **2003**, *68*, 115406.
- [49] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **2008**, *321*, 792.
- [50] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Chem. Rev.* **2012**, *112*, 289.
- [51] W. G. Baber, H. R. Hassé, *Proc. Camb. Phil. Soc.* **1935**, *31*, 564.
- [52] H. Hogreve, *J. Chem. Phys.* **1993**, *98*, 5579.
- [53] A. J. Cohen, P. Mori-Sánchez, *J. Chem. Phys.* **2014**, *140*, 044110.
- [54] C. Kuhn, D. N. Beratan, *J. Phys. Chem.* **1996**, *100*, 10596.
- [55] S. R. Marder, D. N. Beratan, L. T. Cheng, *Science* **1991**, *252*, 103.
- [56] A. Franceschetti, A. Zunger, *Nature* **1999**, *402*, 60.
- [57] T. Weymuth, M. Reiher, *Int. J. Quantum Chem.* **2014**, *114*, 823.
- [58] C. Hermann, J. Neugebauer, M. Reiher, *New J. Chem.* **2007**, *31*, 818.
- [59] M. Wang, X. Hu, D. N. Beratan, W. Yang, *J. Am. Chem. Soc.* **2006**, *128*, 3228.
- [60] S. Keinan, X. Hu, D. Beratan, W. Yang, *J. Phys. Chem.* **2007**, *111*, 176.
- [61] J. H. Werner, S. Kolodinski, H. J. Queisser, *Phys. Rev. Lett.* **1994**, *72*, 3851.
- [62] A. Silverman, A. Zunger, R. Kalish, J. Adler, *Phys. Rev. B* **1995**, *51*, 10795.
- [63] M. Jansen, *Angew. Chem. Int. Ed.* **2002**, *41*, 3746.
- [64] R. P. Feynman, *Phys. Rev.* **1939**, *56*, 340.
- [65] W. Yang, *Phys. Rev. A* **1988**, *38*, 5512.

- [66] S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, *Rev. Mod. Phys.* **2001**, *73*, 515.
- [67] A. Putrino, D. Sebastiani, M. Parrinello, *J. Chem. Phys.* **2000**, *113*, 7102.
- [68] a_r, a_p, b_p, b_r are linear coefficients [eV] for predicting E_p as $= aE_r^{T_l} + b$. Subscripts l and r respectively denote the repulsive wall and attractive tail of the potential. For the four panels in Fig. (3) the parameters are (in same sequence): (a) 0.569, 0.361, -0.137, -0.203; (b) 1.058E-2, 1.473E-4, -7.036E-3, -7.043E-3; (c) 1.069, 0.326, 0.020, -0.197; (d) 5.805E-4, 1.375E-4, -6.920E-3, -0.00692256.
- [69] A. Scherrer, V. Verschinin, D. Sebastiani, *J. Chem. Theory Comput.* **2012**, *8*, 106.
- [70] A. C. Ihrig, A. Scherrer, D. Sebastiani, *J. Chem. Phys.* **2013**, *139*, 094102.
- [71] N. Sablon, F. de Proft, P. W. Ayers, P. Geerlings, *J. Chem. Theory Comput.* **2000**, *6*, 3671.
- [72] S. Fias, Z. Boisdenghien, T. Stuyver, M. Audiffred, G. Merino, P. Geerlings, F. de Proft, *J. Phys. Chem. A* **2013**, *117*, 3556.
- [73] P. Geerlings, F. Fias, Z. Boisdenghen, F. de Proft, *Chem. Soc. Rev.* **2014**, in press.
- [74] D. Benoit, D. Sebastiani, M. Parrinello, *Phys. Rev. Lett.* **2001**, *87*, 226401.
- [75] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.
- [76] W. Andreoni, A. Curioni, *Parallel Computing* **2000**, *26*, 819.
- [77] C. Adamo, M. Cossi, G. Scalmani, V. Barone, *Chem. Phys. Lett.* **1999**, *307*, 265.
- [78] S. Goedecker, M. Teter, J. Hutter, *Phys. Rev. B* **1996**, *54*, 1703.
- [79] C. Hartwigsen, S. Goedecker, J. Hutter, *Phys. Rev. B* **1998**, *58*, 3641.
- [80] A. Willand, Y. O. Kvashnin, L. Genovese, A. Vázquez-Mayagoitia, A. K. Deb, A. Sadeghi, T. Deutsch, S. Goedecker, *J. Chem. Phys.* **2013**, *138*, 104109.
- [81] M. Krack, *Theor. Chim. Acta* **2005**, *114*, 145.
- [82] J. F. Janak, *Phys. Rev. B* **1987**, *18*, 7165.
- [83] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Phys. Rev. B* **2008**, *77*, 155123.
- [84] J. Rezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2011**, *7*, 2427.