Accurate ab initio energy gradients in chemical compound space

O. Anatole von Lilienfeld

Accurate *ab initio* energy gradients in chemical compound space

O. Anatole von Lilienfeld[a)]
Department of Multiscale Dynamic Materials Modeling, Sandia National Laboratories, Albuquerque, New Mexico 87185-1322, USA

Analytical potential energy derivatives, based on the Hellmann–Feynman theorem, are presented for *any* pair of isoelectronic compounds. Since energies are not necessarily monotonic functions between compounds, these derivatives can fail to predict the right trends of the effect of alchemical mutation. However, quantitative estimates without additional self-consistency calculations can be made when the Hellmann–Feynman derivative is multiplied with a linearization coefficient that is obtained from a reference pair of compounds. These results suggest that accurate predictions can be made regarding any molecule's energetic properties as long as energies and gradients of three other molecules have been provided. The linearization coefficent can be interpreted as a quantitative measure of chemical similarity. Presented numerical evidence includes predictions of electronic eigenvalues of saturated and aromatic molecular hydrocarbons. © *2009 American Institute of Physics*. [doi:10.1063/1.3249969]

I. INTRODUCTION

The development of systematic approaches toward the virtual engineering of chemical composition, attempting to computationally identify compound candidates that meet a set of specific desired property requirements, is a worthwhile goal in general, and for pharmaceutical and materials research in particular.[1,2] Due to the large cardinal number of chemical compound space (CCS) already simple enumeration is computationally prohibitive, let alone attempts to parametrize and screen compounds for desired properties. This motivated the development of more "rational" first principles design methodologies[3–5] that usually rely on electronic-structure methods because composition, and thereby chemical bonding, must be varied frequently and freely. Within conventional *ab initio* molecular dynamics (AIMD), an approximate Schrödinger equation is solved and forces obtained for a given compound in order to integrate classical equations of motions for all ionic degrees of freedom.[6,7] Analogously, one would like to extend the set of variables to also include other molecules in order to systematically sample CCS and obtain the desired compound as a statistical mechanical expectation value.

Recently, important steps toward a continuous electronic structure framework of CCS have been made that permit the definition of energetic property gradients with respect to variation in atomic numbers,[8–11] or coefficients in linear combinations of atomic potentials.[12–16] Here, Hellmann–Feynman derivatives[17] in between *any* pair of isoelectronic compounds are presented and numerically evaluated for an energy difference, the highest occupied molecular orbital (HOMO) eigenvalue. Unfortunately, due to the nonmonotonous behavior of the HOMO eigenvalue in between two compounds, the derivative can fail even qualitatively to esti-

mate the effect of compositional variation. Based on a linearizing coefficient, an *Ansatz* is developed in this study that circumvents this problem and that drastically improves the ability to quantitatively predict changes due to alchemical mutations without having to perform any self-consistency calculation for the mutant. The presented results therefore hold promise to become valuable for AIMD sampling of CCS, for efficient gradient-based optimization algorithms in CCS, or for more efficient computation of free energy differences through thermodynamic integration over alchemical changes.[18–23]

II. METHODOLOGY

A. Transmutation

As frequently exploited for statistical mechanical computation of free energy differences[21] between compounds based on thermodynamic integration,[18] or free-energy perturbations,[24] any state function can reversibly be connected through an order parameter, $0 \leq \lambda \leq 1$, usually through linear interpolation of the two Hamiltonians. For the potential energy, E, of any pair of compounds i and j,

$$E(\lambda) = E[H_{ij}(\lambda)] = E[H_i + \lambda \cdot (H_j - H_i)]. \tag{1}$$

Since the electronic unperturbed Hamiltonian can only differ in number of electrons, N_e, and nuclear charge distribution, $Z(\mathbf{r}) = \Sigma_I N_I \delta(\mathbf{r} - \mathbf{R}_I)$ (N_I and $\mathbf{R}_I$ being the atomic number and position of atom I),[9] this interpolation can be performed in the plane spanned by N_e and $Z(\mathbf{r})$ (see Fig. 1).

Changes in N_e have already been used to study redox processes in gas, biological, and liquid systems,[9,23,25] based on conceptual DFT and the energy derivative corresponding to the HOMO eigenvalue within Kohn–Sham DFT.[26–31] Dealing exclusively with isoelectronic changes in $Z(\mathbf{r})$, however, derivatives can naturally be defined through linear combination of "alchemical potentials" evaluated at all the nuclear positions, according to Refs. 8–11. This approach,

---

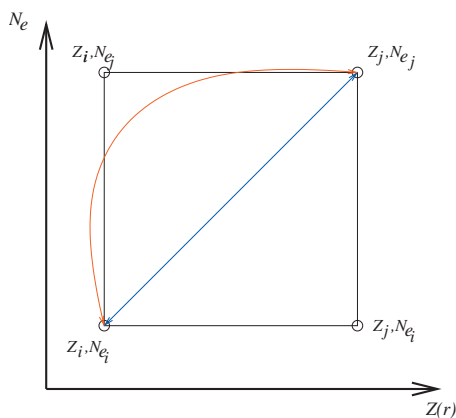a)Electronic mail: oavonli@sandia.gov. URL: http://www.sandia.gov/~oavonli.

FIG. 1. Cartoon of alchemical transform plane between compounds i and j with set of electron number, N_e, and high-dimensional nuclear charge distribution, $Z(\mathbf{r})$, according to Eq. (1). The blue diagonal represents the linearly interpolated Hamiltonian, the red line one of the manifold of nonlinear ones, both driven by a coupling parameter, $0 \le \lambda \le 1$. This paper deals exclusively with horizontal, i.e., isoelectronic, changes and derivatives, $N_{e_i} = N_{e_j}$, $E(\lambda)|_{N_e}$, and $\partial_\lambda E|_{N_e}$.

FIG. 2. Linear transmutation of HOMO eigenvalue, $\epsilon_{ij} = \epsilon[H_{ij}(\lambda)]$, for relaxed and superimposed compound pairs, $(i, j \in CH_4, NH_3, H_2O, HF)$. The inset (same symbol/color code) demonstrates the good numerical agreement between finite difference derivative (symbol) and analytical Hellmann–Feynman derivative (star) using v_{ij}^t as displayed in Fig. 3 and according to Eq. (5).

however, is hampered by the fact that such derivatives refer to paths which follow infinitesimal changes in the classical nuclear charge distribution, $Z(\mathbf{r}) \mapsto Z(\mathbf{r}) + \delta Z$, implying that for every nucleus one must sequentially pass through adjacent atomic numbers in the periodic table.

For paths restricted to *any* two isoelectronic compounds, however, it is equally straightforward to leverage the Hellmann–Feynman theorem,

$$\frac{dE[H_{ij}(\lambda)]}{d\lambda} = \left\langle \frac{\partial H_{ij}(\lambda)}{\partial \lambda} \right\rangle_\lambda = \int d\mathbf{r} n_\lambda(\mathbf{r}) \cdot v_{ij}^t(\mathbf{r}), \quad (2)$$

where $\langle \cdots \rangle$ denotes the quantum-mechanical expectation value of the local derivative of the transmutating Hamiltonian, dubbed the "transmutational" potential, $v_{ij}^t(\mathbf{r})$. For the linearly interpolated Hamiltonian, Eq. (2) amounts to the expectation value of the difference between the external potentials of compounds i and j,

$$\frac{dE(\lambda)}{d\lambda} = \langle H_j - H_i \rangle_\lambda = \int d\mathbf{r} n_\lambda(\mathbf{r}) \cdot [v_j(\mathbf{r}) - v_i(\mathbf{r})]. \quad (3)$$

Here, $v_i(\mathbf{r})$ refers to the conventional Coulomb potential of the nuclear charges acting on electrons, $\Sigma_{I \in i} N_I / |\mathbf{r} - \mathbf{R}_I|$, the trivial contribution from the internuclear Coulomb repulsion being omitted for simplicity. Note how the λ dependence of $\partial_\lambda E(\lambda)$ is introduced exclusively by the electron density $n(\mathbf{r})$, v^t being a constant object. This is to be contrasted with the discontinuous stepwise behavior of $\partial_N E(N_e)|_{Z(\mathbf{r})}$.[28]

The restriction to being isoelectronic, on a side note, is by no means serious. The diversity of CCS is rather due to all the possible specific combinations of functional groups in real space (local objects), than due to the number of electrons of the entire system (global object). Furthermore, in most practical applications of property optimization, the number of electrons is quite finite due to previously selected classes of compound systems. In the case of drug design, for example, maybe dozens of different numbers of electrons might have to be screened, but not hundreds or thousands. Finally, the isoelectronic requirement becomes even less se-
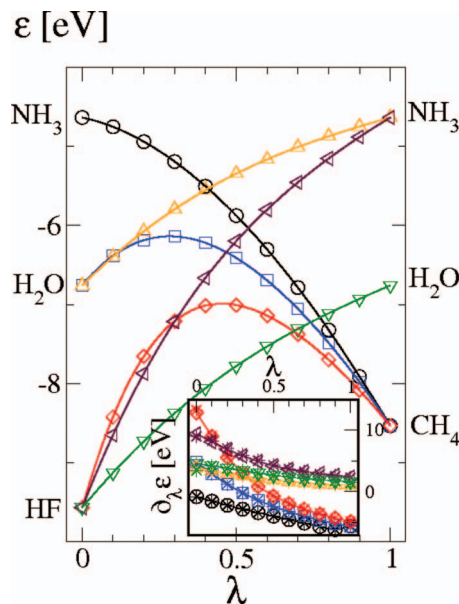
vere if, as it is shown to be the case within pseudopotential calculations, this restriction refers only to the valence-electron number. If one considers, for example, all the possible isoelectronic molecules containing only 30 valence electrons (benzene just as well as $B_3N_3H_6$ just as well as Si_6H_6), in particular when including all rows (through pseudopotentials) and columns (through satisfying valency by adding or removing hydrogen). In practice, for a given energetic property optimization one could easily envision to perform separate isoelectronic gradient based optimization for a small set of different electron numbers.

Linear isoelectronic transmutations have been computed for all compound pairs of methane, ammonia, water, and hydrogen fluoride for a potential energy difference, namely, the HOMO eigenvalue,

$$\epsilon_{ij}(\lambda) \approx \frac{E[H_{ij}(N_e, \lambda)] - E[H_{ij}(N_e - \delta, \lambda)]}{\delta}, \quad (4)$$

for small δ. The results (Fig. 2) indicate a smooth, concave, and overall well-behaved dependence. The Hellmann–Feynman derivative, computed by combination of Eqs. (3) and (4),[27]

$$\frac{d\epsilon_{ij}(\lambda)}{d\lambda} = \frac{1}{\delta} \int d\mathbf{r} [n_\lambda(\mathbf{r}) - n_\lambda^{\delta+}(\mathbf{r})] \cdot [v_j(\mathbf{r}) - v_i(\mathbf{r})], \quad (5)$$

reflects this behavior perfectly well. Numerical results for these derivatives [Eq. (5)] as a function of λ are displayed as an inset for all compound pairs in Fig. 2, together with the corresponding coinciding finite difference derivatives, $d\epsilon / d\lambda \approx [\epsilon(\lambda) - \epsilon(\lambda - \delta)] / \delta$. The very good agreement between these derivatives serves as numerical evidence for the suitability of the Hellmann–Feynman theorem in this context. Illustrative isosurfaces of the employed transmutational
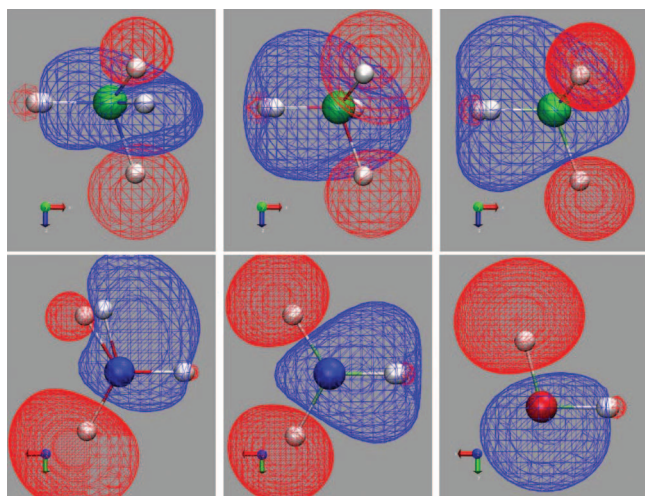
FIG. 3. Transmutational potentials, $v_{ij}^t(\mathbf{r}) = v_j(\mathbf{r}) - v_i(\mathbf{r})$, used for computation of energy derivatives in Figs. 2 and 4 according to Eq. (2). Six linearly interpolated, relaxed, and superimposed compound pairs (i,j). (From left to right) Top: (CH_4, NH_3), (CH_4, H_2O), and (CH_4, HF). Bottom: (NH_3, H_2O), (NH_3, HF), and (H_2O, HF). Blue and red represent isovalue surfaces at 0.2 and -0.2 a.u., respectively.

potentials, $v_j(\mathbf{r}) - v_i(\mathbf{r})$, which have been used to compute the analytical derivatives, are featured in Fig. 3. The chemical nature of both molecules is clearly encoded in this object, evoking an interpretation of Eqs. (2) and (3) as alchemical perturbational response.

B. Linearizing

Inspection of the connecting curves makes clear that these gradients are not necessarily well suited, and in the nonmonotonic cases can fail to predict even qualitatively the right trends due to transmutation. Suppose, for example, one desired to predict the HOMO eigenvalue of methane after having calculated the HOMO eigenvalue of water (blue line in Fig. 2). With the transmutational analytical derivative, requiring only the external potential of methane as an input, one could attempt to predict methane's value by truncating a Taylor expansion to first degree in λ,

$$\epsilon_{CH_4} \approx \epsilon_{H_2O} + \left.\frac{\partial \epsilon_{H_2O, CH_4}}{\partial \lambda}\right|_{\lambda \equiv H_2O} \cdot \Delta\lambda, \quad \Delta\lambda = +1. \quad (6)$$

While the derivative at $\lambda = 0$ (corresponding to H_2O) is correctly pointing upward (solid arrow), the actual HOMO eigenvalue of CH_4 is below the one of H_2O, thereby yielding a qualitatively wrong prediction of the effect of transmutation. Resorting to higher order of the property expansion in λ, based on linear response theory for example,[32] one might hope to estimate at least qualitatively the right trends. The computational burden, however, would also increase, thereby defying the initial motive of being able to predict the effect of the transmutation analytically, i.e., without additional computational overhead.

If, however, the energy (difference), and not its Hamiltonian, was linear in λ, the first order derivative would be sufficiently predictive by construction. Since the interpolation of the Hamiltonian is arbitrary, instead of Eq. (1), one can equally well postulate a new Hamiltonian with energy,

$$E^L(\lambda) = E[H_{ij}^L(\lambda)] = E[H_i + f_{ij}(\lambda) \cdot (H_j - H_i)], \quad (7)$$

where f_{ij} is a function with boundary conditions $f_{ij}(\lambda=0) = 0$ and $f_{ij}(\lambda=1) = 1$, whose sole purpose consists of linearizing E in λ. Assuming f to be harmonic, the boundary conditions define all but one coefficient, a_{ij}, and

$$H_{ij}^L(\lambda) = H_i + (\lambda + a_{ij} \cdot \lambda \cdot [1 - \lambda]) \cdot (H_j - H_i), \quad (8)$$

a result that is reminiscent of the enthalpy of a binary mixture in regular solution theory. Combining Eqs. (2) and (8), equating them to the new constraint that E be linear in λ, i.e., $\partial_\lambda E^L = E_j - E_i$, one solves for a_{ij},

$$a_{ij}(\lambda) = \left(\frac{E_j - E_i}{\partial_\lambda E_{ij}(\lambda)} - 1\right) \cdot \frac{1}{2\lambda - 1}. \quad (9)$$

Note the divergence for $\lambda \to 1/2$, and that if, by coincidence, $E(\lambda)$ was already linear for the linearly interpolated H, a_{ij} would be zero. a_{ij} depends not only explicitly on λ but also implicitly through the electron density. At the two end points, however, H_{ij} does not depend on λ, and a_{ij} is given analytically, $a_{ij}(\lambda=0) = 1 - [E_j - E_i]/\partial_\lambda E_{ij}(\lambda=0)$, and $a_{ij}(\lambda=1) = [E_j - E_i]/\partial_\lambda E_{ij}(\lambda=1)$. For any λ, a_{ij} requires knowledge about both end points.

C. Linearizing coefficients

What if, however, a_{ij} was transferable, i.e., one could determine it once for a reference compound pair, (i', j'), and then use it to linearize other pairs of compounds (i, j)? Insertion of $a_{i'j'}^{ref}$ into Eq. (8) and differentiation according to Eq. (2) yields such an approximation to the exact first order derivative of $E^L(\lambda)$ at $\lambda = 1$ or 0, using information of only either of the two end points,

$$\partial_\lambda E^L(\lambda) \approx \left(\frac{\partial_\lambda E_{i'j'}^L}{\partial_\lambda E_{i'j'}(\lambda)}\right)^{ref} \cdot \partial_\lambda E_{ij}(\lambda). \quad (10)$$

The fraction, expressing the degree of divergence in slope between E^L and E in the case of the reference pair, stems from $a_{i'j'}^{ref}$, and is dubbed "linearization coefficient." Note that the numerator, $\partial_\lambda E_{i'j'}^L$, is identical with $E_{j'} - E_{i'}$.

In order to test this *Ansatz* HOMO eigenvalues have been estimated at one end point ($\lambda = 1$) using eigenvalue, gradients, and C^{ref} at the other end point ($\lambda = 0$),

$$\epsilon_{ij}(\lambda = 1) \approx \epsilon_{ij}(\lambda = 0) + C_{i'j'}^{ref}(\lambda = 0) \cdot \partial_\lambda \epsilon_{ij}(\lambda = 0) \cdot \Delta\lambda,$$

$$\text{with} \quad C_{i'j'}^{ref}(\lambda) = \left(\frac{\epsilon_{j'} - \epsilon_{i'}}{\partial_\lambda \epsilon_{i'j'}(\lambda)}\right)^{ref}, \quad (11)$$

the linearization coefficient, C^{ref}, being defined in complete analogy to Eq. (10). For numerical evaluation, (i,j) and (i', j') have been restricted to compound pairs involving the prediction of molecules with one functional group, G_i, starting from the same molecules but with a different functional group, G_j. Referencing has been included through variation of the remainder of the molecule, R, i.e., (i,j) and (i', j') have been modeled by $(R\text{-}G_i, R\text{-}G_j)$ and $(R'\text{-}G_i, R'\text{-}G_j)$, respectively. More specifically for $(G_i = CH_3)$,
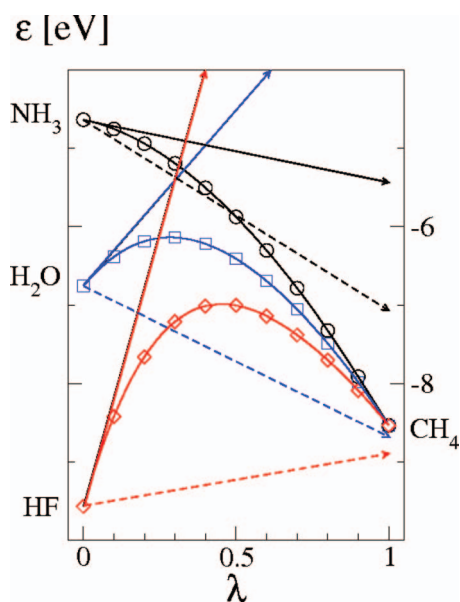
FIG. 4. Prediction (arrows) of HOMO eigenvalue of CH$_4$ (-8.54 eV) from NH$_3$, H$_2$O, and HF, according to Eqs. (11) and (12). Solid arrows correspond to predictions based solely on the first order derivative of ϵ_{ij}, [Eq. (6)]. Dashed arrows use linearization coefficients, C^{ref}, as obtained from reference pairs, $(R'\text{-}G_i, R'\text{-}G_j) \in (\text{CH}_3\text{--CH}_3, \text{CH}_3\text{--NH}_3)$, $(\text{CH}_3\text{--CH}_3, \text{CH}_3\text{--OH})$, and $(\text{CH}_3\text{--CH}_3, \text{CH}_3\text{--F})$ [Eq. (12)]. Corresponding numerical results for this and other cases are presented in Table I.

$$\epsilon^{\text{pred}}_{R\text{-CH}_3} = \epsilon^{\text{start}}_{R\text{-}G_j} + C^{\text{ref}}_{R'\text{-CH}_3, R'\text{-}G_j} \cdot \partial_\lambda \epsilon^{\text{start}}_{R\text{-}G_j}. \tag{12}$$

A graphic illustration of the increase in predictive power using the linearizing coefficients is given in Fig. 4 (dashed arrows) for R=H, $G_j \in$ NH$_2$, OH, F, and R'=CH$_3$, i.e.,

$$\epsilon^{\text{pred}}_{\text{CH}_4} \approx \epsilon^{\text{start}}_{\text{NH}_3} + C^{\text{ref}}_{\text{CH}_3\text{--CH}_3, \text{CH}_3\text{--NH}_2} \bigg| \frac{\partial \epsilon^{\text{start}}_{\text{NH}_3}}{\partial \lambda}, \tag{13}$$

$$\approx \epsilon^{\text{start}}_{\text{H}_2\text{O}} + C^{\text{ref}}_{\text{CH}_3\text{--CH}_3, \text{CH}_3\text{--OH}} \cdot \frac{\partial \epsilon^{\text{start}}_{\text{H}_2\text{O}}}{\partial \lambda}, \tag{14}$$

$$\approx \epsilon^{\text{start}}_{\text{HF}} + C^{\text{ref}}_{\text{CH}_3\text{--CH}_3, \text{CH}_3\text{--F}} \cdot \frac{\partial \epsilon^{\text{start}}_{\text{HF}}}{\partial \lambda}. \tag{15}$$

This figure is discussed in more detail in Sec. III.

Further predictions of $\epsilon^{\text{pred}}_{R\text{-CH}_3}$ have been calculated starting from $(G_j \in \text{NH}_2, \text{OH}, \text{F})$, for $[R, R' \in \text{H}, \text{CH}_3, \text{C}_2\text{H}_5, \text{Ph}$ (benzene ring: C$_6$H$_5$), Py (*para*-substituted pyridine ring: p-NC$_5$H$_4$)]. For all pairs $(R\text{-CH}_3, R\text{-}G_j)$, the geometries of the various R have been kept constant, thereby canceling the nuclear contribution of R to the derivative. All transmutating groups, (CH_3, G_j), have been relaxed before being superimposed at the position of the heavy atom, and attached to R.

D. Computational details

All calculations have been performed within Kohn–Sham density functional theory,[26,33] using the generalized gradient approximation PBE,[34] as implemented in the plane-wave pseudopotential code CPMD,[35] Goedecker pseudopotentials[36] as published by Krack,[37] a plane-wave cutoff of 200 Ry, isolated box of 15^3 Å^3 for CH$_4$-series,

$20 \times 15 \times 15$ Å^3 for C$_2$H$_6$ and C$_3$H$_8$ series, and $20 \times 20 \times 10$ Å^3 for Ph and Py series. The transmutational potential, $v^t = v_j - v_i$, as depicted in Fig. 3 using VMD (Ref. 38) has been computed by superimposing the position of the heavy atoms which are being transformed, and by subtracting the total potentials due to the local pseudopotentials of all atoms involved in the transmutation. A routine for the calculation of this expectation value has been implemented in CPMD. The Hellmann–Feynman derivatives of the HOMO eigenvalues have been obtained via finite difference for a slight decrease in N_e by δ through introduction of a background charge of δ=+0.01. For simplification, and without any loss of generality, the nonlocal channel of the fluorine atom has been used for the transforming heavy atoms, and kept constant throughout all transmutations. While this affects the absolute values of the real corresponding PBE HOMO eigenvalues, ϵ_{CH_4} =-9.26, $\epsilon_{\text{C}_2\text{H}_6}$=$-8.05$, $\epsilon_{\text{C}_3\text{H}_8}$=$-7.71$, $\epsilon_{\text{Ph-CH}_3}$=$-5.80$, and $\epsilon_{\text{Py-CH}_3}$=$-5.48$ eV, their ranking is not altered. Calculations for noninteger λ values have been performed by scaling the valence atomic number and pseudopotential parameters $\{C_i\}$ in the pseudopotentials. The geometries of the benzene and the pyridine group correspond to the relaxed geometry of benzene. The heavy atom of the transmutating group has always been placed at 1.511 Å distance from the aromatic ring. All hydrogen atoms of the transmutating group have been relaxed for the benzene system, except for molecules in the CH$_4$ series which have been relaxed separately.

III. RESULTS AND DISCUSSION

An illustrative example of the increase in predictive accuracy is given by the arrows in Fig. 4 for $(G_i=\text{CH}_3)$, $(G_j \in \text{NH}_2, \text{OH}, \text{F})$, $(R=\text{H})$, and $(R'=\text{CH}_3)$. Clearly, the prediction based solely on the Hellmann–Feynman derivative is rather meaningless: NH$_3$, H$_2$O, and HF predict methane's HOMO eigenvalue (-8.54 eV) to be -5.43, -2.28, and $+4.41$ eV, respectively. Only after multiplication with the reference coefficient, quantitative results can be obtained: -7.07, -8.68, and -8.90 eV for NH$_3$, H$_2$O, and HF using $R'=\text{CH}_3$ for the reference pair.

All the corresponding numerical results for this and all other combinations of $\{R, R'\}$ are displayed in Table I. One finds a systematic and sometimes drastic improvement in predictive accuracy when using the reference coefficient. The root-mean-square deviation from the actual eigenvalue is decreased to eV accuracy, no matter how large it is for predictions without referencing. As one would expect for the various starting groups, $\{G_j\}$, the improvement is more significant for cases with the more pronounced extrema in the connecting curves, $-$F$>$$-OH>$$-NH_2$. Furthermore, it seems that the larger R the less dramatic is the deviation of the unreferenced prediction from the actual value. If the improvement is a measure of similarity of compound pairs, these results suggest that linearizing reference coefficients might offer a general way to quantify otherwise ill-posed notions such as molecular similarity or transferability. For example, the ranking in accuracy of the prediction of the HOMO eigenvalue of propane, C$_2$H$_5$$-CH_3$, starting from C$_2H_5$$-$F follows $C^{\text{ref}}_{\text{CH}_3\text{-CH}_3,\text{CH}_3\text{-F}} > C^{\text{ref}}_{\text{H-CH}_3,\text{H-F}} > C^{\text{ref}}_{\text{Ph-CH}_3,\text{Ph-F}}$

TABLE I. Predicted HOMO eigenvalues, $\epsilon_{R\text{-}CH_3}^{pred}$, of various molecules $\{R\text{-}CH_3\}$. All predictions made starting from $R\text{-}G_j$ with $(G_j \in F, OH, NH_2)$ and according to Eq. (12). Diagonal values (bold) refer to predictions made without referencing $(C_{R'\text{-}G_j, R'\text{-}CH_3}^{ref} = 1)$, off-diagonal predictions correspond to the diagonal predictions multiplied by $\{C_{R'\text{-}F, R'\text{-}CH_3}^{ref}\}$, $\{C_{R'\text{-}OH, R'\text{-}CH_3}^{ref}\}$, or $\{C_{R'\text{-}NH_2, R'\text{-}CH_3}^{ref}\}$ in the respective top, middle, and lower third of this table $(C_{R'\text{-}G_j, R'\text{-}CH_3}^{ref} = (\epsilon_{R'\text{-}CH_3} - \epsilon_{R'\text{-}G_j}) / \partial_\lambda \epsilon_{R'\text{-}G_j})$. First and second row entries in the $H\text{-}CH_3$-column correspond to solid and dashed arrows in Fig. 2 (right). The actual HOMO eigenvalues $\epsilon_{R\text{-}CH_3}^{act}$ and their root-mean-square deviations are shown in the last three rows, with $(C_{R'\text{-}G_j, R'\text{-}CH_3}^{ref} = (\epsilon_{R'\text{-}CH_3} - \epsilon_{R'\text{-}G_j}) / \partial_\lambda \epsilon_{R'\text{-}G_j})$ and without $(C_{R'\text{-}G_j, R'\text{-}CH_3}^{ref} = 1)$ (bold) linearizing coefficient. All values are in eV.

Starting $R'\text{-}G_j$	$R\text{-}CH_3$				
	$H-CH_3$	CH_3-CH_3	$C_2H_5-CH_3$	$Ph-CH_3$	$Py-CH_3$
H-F	**+4.41**	−7.37	−7.25	−6.05	−5.90
CH_3−F	−8.90	**+2.15**	−7.45	−6.14	−5.95
C_2H_5−F	−9.01	−7.72	**+0.06**	−6.17	−5.97
Ph-F	−7.65	−6.72	−6.75	**−2.75**	−5.76
Py-F	−6.55	−5.91	−6.13	−5.54	**−3.93**
H-OH	**−2.28**	−7.51	−7.27	−6.37	−6.17
CH_3−OH	−8.68	**−1.94**	−7.38	−6.44	−6.22
C_2H_5−OH	−8.85	−7.79	**−2.31**	−6.53	−6.26
Ph-OH	−7.49	−6.58	−6.44	**−3.10**	−5.87
Py-OH	−6.51	−5.71	−5.67	−5.31	**−4.38**
$H-NH_2$	**−5.43**	−9.87	−9.95	−5.86	−6.22
CH_3-NH_2	−7.07	**−5.15**	−7.71	−5.27	−5.71
$C_2H_5-NH_2$	−6.94	−7.45	**−5.20**	−5.22	−5.67
$Ph-NH_2$	−6.30	−9.75	−9.84	**−4.63**	−6.20
$Py-NH_2$	−6.74	−7.14	−7.20	−5.14	**−5.14**
$\epsilon_{R\text{-}CH_3}^{act}$	−8.54	−7.64	−7.52	−5.82	−5.59
RMSD[a]	**8.50**	**6.70**	**5.48**	**2.47**	**1.21**
RMSD[b]	1.40	1.24	1.25	0.49	0.45

[a]$C_{R'\text{-}G_j, R'\text{-}CH_3}^{ref} = 1$.
[b]$C_{R'\text{-}G_j, R'\text{-}CH_3}^{ref} = (\epsilon_{R'\text{-}CH_3} - \epsilon_{R'\text{-}G_j}) / \partial_\lambda \epsilon_{R'\text{-}G_j}$.

$> C_{Py\text{-}CH_3, Py\text{-}F}^{ref}$, thereby providing a specific and quantitative ranking of similarity, in terms of HOMO eigenvalues, among these compound pairs. Predictions of compounds with $R\text{-}G_i$ other than $R\text{-}CH_3$, namely, for $G_i \in NH_2$, OH, or F (not shown in this study), also improve when using linearization coefficients, albeit less dramatically.

IV. CONCLUSIONS

It has been shown that analytical gradients of potential energies in CCS can be obtained between any isoelectronic pair of compounds according to the Hellman–Feynman theorem. Sets of linearization coefficients can be employed to drastically improve the accuracy of predicting changes in HOMO eigenvalues due to alchemical transmutation without any significant increase in computational effort. Albeit requiring preliminary effort for establishing a library of compound pairs, this *Ansatz* is not only promising for the development of more efficient gradient-based optimization algorithms navigating high-dimensional subsets of CCS, it might also prove useful for more efficient free energy difference evaluations relying on numerical thermodynamic integration.[23] Furthermore, linearization coefficients might offer a useful interpretative means to quantify chemical similarity.

The submitted study represents important progress with respect to variational changes in nuclear charge distribution.[9]

Through generalization of the derivatives as expectation values of differences in external potentials any two isoelectronic compounds (no matter how different the atomic numbers of their constituting atoms, no matter how different the number and location of their atomic sites) can now be directly connected and their energy derivatives be obtained analytically simply by virtue of the Hellmann–Feynman theorem. Furthermore, the problem of nonmonotonous functions in between two compounds is circumvented via the linearizing reference coefficients. Hence, it seems possible to rigorously and quantitatively predict one molecule's energetic properties based on given knowledge about one isoelectronic molecule, augmented by library information of another molecule pair.

It is reiterated that while this study has been limited to atoms of the first and second row only, the restriction to isoelectronic interpolations is sufficiently general to permit the extension to elements of other rows in the periodic table simply through the use of pseudopotentials, thereby factorially increasing CCS, i.e., the number of available compound pairs and corresponding derivatives. The technical details of numerical pseudopotential interpolation in conjunction with known effects of pseudopotential construction on relevant electronic properties, such as band gaps,[39] will be part of future studies. The suitability of this approach for dealing

with other systems and other properties to which the Hellmann–Feynman theorem does not apply remains to be elucidated.

[1] P. Kirkpatrick and C. Ellis, Nature (London) **432**, 823 (2004).
[2] J. Hafner, C. Wolverton, G. Ceder, and Guest Editors, MRS Bull. **31**, 659 (2006).
[3] A. Franceschetti and A. Zunger, Nature (London) **402**, 60 (1999).
[4] P. Piquini, P. A. Graf, and A. Zunger, Phys. Rev. Lett. **100**, 186403 (2008).
[5] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Nat. Chem. **1**, 37 (2009).
[6] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[7] R. Iftimie, P. Minary, and M. E. Tuckerman, Proc. Natl. Acad. Sci. U.S.A. **102**, 6654 (2005).
[8] O. A. von Lilienfeld, R. Lins, and U. Rothlisberger, Phys. Rev. Lett. **95**, 153002 (2005).
[9] O. A. von Lilienfeld and M. E. Tuckerman, J. Chem. Phys. **125**, 154104 (2006).
[10] O. A. von Lilienfeld and M. E. Tuckerman, J. Chem. Theory Comput. **3**, 1083 (2007).
[11] V. Marcon, O. A. von Lilienfeld, and D. Andrienko, J. Chem. Phys. **127**, 064305 (2007).
[12] M. Wang, X. Hu, D. N. Beratan, and W. Yang, J. Am. Chem. Soc. **128**, 3228 (2006).
[13] D. Xiao, W. Yang, and D. N. Beratan, J. Chem. Phys. **129**, 044106 (2008).
[14] X. Hu, D. N. Beratan, and W. Yang, J. Chem. Phys. **129**, 064102 (2008).
[15] D. Balamurugan, W. Yang, and D. N. Beratan, J. Chem. Phys. **129**, 174105 (2008).
[16] S. Keinan, M. J. Therien, D. N. Beratan, and W. Yang, J. Phys. Chem. A **112**, 12203 (2008).
[17] R. P. Feynman, Phys. Rev. **56**, 340 (1939).
[18] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[19] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[20] B. Widom, J. Chem. Phys. **39**, 2808 (1963).
[21] W. F. van Gunsteren, X. Daura, and A. E. Mark, Helv. Chim. Acta **85**, 3113 (2002).
[22] M. Sulpizi and M. Sprik, Phys. Chem. Chem. Phys. **10**, 5238 (2008).
[23] K. Leung, S. B. Rempe, and O. A. von Lilienfeld, J. Chem. Phys. **130**, 204507 (2009).
[24] C. Oostenbrink, J. Comput. Chem. **30**, 212 (2009).
[25] X. Zeng, H. Hu, X. Hu, A. J. Cohen, and W. Yang, J. Chem. Phys. **128**, 124510 (2008).
[26] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[27] J. F. Janak, Phys. Rev. B **18**, 7165 (1978).
[28] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).
[29] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
[30] P. Geerlings, F. D. Proft, and W. Langenaeker, Chem. Rev. (Washington, D.C.) **103**, 1793 (2003).
[31] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Phys. Rev. B **77**, 115123 (2008).
[32] A. Putrino, D. Sebastiani, and M. Parrinello, J. Chem. Phys. **113**, 7102 (2000).
[33] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[34] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[35] J. Hutter *et al.*, computer code CPMD, V3.13.2, Copyright IBM Corp. 1990–2008, Copyright MPI-FKF Stuttgart 1997–2001.
[36] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
[37] M. Krack, Theor. Chim. Acta **114**, 145 (2005).
[38] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[39] O. A. von Lilienfeld and P. Schultz, Phys. Rev. B **77**, 115202 (2008).