

Molecular grand-canonical ensemble density functional theory and exploration of chemical space

O. Anatole von Lilienfeld^{a)}

Department of Chemistry, New York University, New York, New York 10003 and Institute for Pure and Applied Mathematics, University of California Los Angeles, Los Angeles, California 90095

Mark E. Tuckerman

Department of Chemistry, New York University, New York, New York 10003; Institute for Pure and Applied Mathematics, University of California Los Angeles, Los Angeles, California 90095; and Courant Institute of Mathematical Sciences, New York University, New York, New York 10003

(Received 23 May 2006; accepted 28 July 2006; published online 18 October 2006)

We present a rigorous description of chemical space within a *molecular* grand-canonical ensemble multi-component density functional theory framework. A total energy density functional for chemical compounds in contact with an electron and a proton bath is introduced using Lagrange multipliers which correspond to the energetic response to changes of the elementary particle densities. From a generalized Gibbs-Duhem equation analog, reactivity indices such as the *nuclear* hardness and a *molecular* Fukui function, which couples the grand-canonical electronic and nuclear degrees of freedom, are obtained. Maxwell relations between composition particles, ionic displacements, and the external potential are discussed. Numerical results for the molecular Fukui function are presented as well as finite temperature estimates for the oxidation of ammonia. © 2006 American Institute of Physics. [DOI: 10.1063/1.2338537]

I. INTRODUCTION

Designing molecules with optimized properties, an aspect of “reverse engineering” also known as *rational compound design*,^{1–5} is best understood via the concept of *chemical space*, a notion that is receiving increased attention.^{6,7} In this paper, we aim to establish the underpinnings for linking the idea of chemical space to a *molecular* grand-canonical ensemble (GCE) theory within the electronic density functional theory (DFT) framework. The potential utility and power of a molecular GCE theory can be exemplified when considering the problem of maximizing the relative binding free energy of a drug to a target receptor as the chemical composition of a functional group is varied.

While chemical space can be defined in various ways, the working definition that will be employed here is the discrete molecular space spanned by all stable chemical compounds. This seemingly simple definition contains numerous subtleties, however, that require further elaboration. We thus begin with a fundamental description of matter and build up the concept of chemical space upon which the present theory and applications will be based. This discussion will rely on the concept of a λ path, wherein a switching parameter, λ , allows a system to be smoothly transformed from one discrete thermodynamic chemical state into another via an arbitrary, i.e., not necessarily realistic, path.

This paper is organized as follows: First, in Sec. II, an overview is given which outlines the most important underlying concepts, assumptions, results, and conclusions of this study. Then, in Sec. III, we present the theoretical framework

for a molecular grand-canonical ensemble electronic DFT. In Sec. IV, two numerical applications are discussed. Finally, conclusions are given in Sec. V.

II. OVERVIEW

A. Chemical space

Ignoring the substructure of nucleons, the most fundamental description of matter in terms of elementary particles requires specification of the number of electrons (N_e), protons (N_p), and neutrons (N_n) in the system and, therefore, a “point” in the chemical space at this level is simply any stable ordered triple (N_e, N_p, N_n) . A full quantum field-theoretical treatment of all of the electromagnetic and nuclear interactions would, at least in principle, describe the processes by which one stable compound formed from this combination of particles transforms into another stable compound. For example, the choice $N_e=N_p=N_n=10$ could form Ne or D₂O. However, a process by which two protons and two neutrons escape from the Ne nucleus and capture two electrons to form the deuterium and the oxygen atoms never occurs under normal conditions and, in any case, is irrelevant for practical chemical applications. Nevertheless, this type of treatment would render essentially moot the notion of its associated molecular chemical space since all compounds for a given (N_e, N_p, N_n) could be described through an ergodic sampling of the real pathways provided within such a fundamental description. Crudely speaking, a maximal number of “physical” (canonical) degrees of freedom renders the number of “chemical” (grand-canonical) degrees of freedom minimal.

For quantum-chemical applications, however, a more

^{a)}Electronic mail: ovt203@nyu.edu. URL: <http://homepages.nyu.edu/~ovt203/>

useful description entails “bundling” the protons and neutrons together into the nuclei of the known chemical elements, neglecting the underlying nuclear physics, and using a standard nonrelativistic electronic Hamiltonian. At this level, the only role played by the neutrons is to allow different isotopes to be included in the treatment. Although potentially important in some applications, for simplicity we shall restrict ourselves to the “standard” (nonisotopic) chemical elements, thereby allowing us to ignore the N_n degree of freedom in chemical space and treat only N_p , which is related to the sum over all nuclear core charges, and N_e the corresponding number of electrons. While a full sampling of such a chemical space requires full variation in N_p and N_e , usually chemical stability implies that $N_e \approx N_p$. At this quantum-chemical level of treatment, an additional degree of freedom is needed to specify the bundling, i.e., the number of stable nuclei that can be formed for each choice of N_p protons. In a system with $N_e = N_p = 10$, for example, the systems (He)₅, CH₄, NH₃, H₂O, HF, and Ne all represent distinct points in this chemical space. The only way to sample these different compounds is through an unrealistic path through the chemical space, e.g., using a λ -switching scheme. Since our interest will exclusively be in the calculation of thermodynamic state functions such as the free energy, arbitrary λ paths are perfectly valid and highly useful. From this description, it becomes clear that restricting the number of physical degrees of freedom, i.e., by restricting the protons to be bundled into the known chemical elements, the number of degrees of freedom in the chemical space necessarily increases. Notice, however, that the description at this level still assumes that for each choice of the atomic identities, different chemical bonding arrangements and different conformers can be sampled on a physical potential energy surface rather than through λ paths, thereby implying that the explicit specification of N_e and N_p requires an electronic structure method treatment of the system, wherein such an assumption is valid.

However, as the preceding discussion makes clear, further physical degrees of freedom can be “interchanged” with chemical-space degrees of freedom. Thus, for example, when an empirical force field is employed, the chemical bonding pattern is usually assumed *a priori* and is held static, which means that the N_e degree of freedom is rendered moot and that the chemical space acquires a new degree of freedom, namely, the number of chemical bonding patterns possible for a given set of atoms. As already noted, these arrangements can then be sampled via λ paths in the chemical space. Interestingly, even at an *ab initio* level, it can prove computationally advantageous to include the choice to visit different chemical bonding arrangements either via enhanced sampling of a set of realistic reaction coordinates or via a λ -switching path in the chemical space—both involving the corresponding electronic and ionic rearrangements.

Finally, one can consider the extreme limit corresponding to a molecule at very low or zero temperature, in which different conformers are not sampled by overcoming the realistic barriers on the potential energy surface but rather by including them in the chemical space and accessing them *only* via unrealistic λ paths. This would be the limit in which the number of real degrees of freedom is minimal, only cor-

responding to the vibrational zero-point modes, while the number of chemical-space degrees of freedom becomes exponentially large. As an example, a simple alkane chain C_nH_{2n+2}, with three conformational minima per dihedral angle, would span a chemical space which includes 3^{n-3} rotamers.

From this discussion, one can conclude that for a quantum-chemical treatment of chemical space, the two fundamental parameters N_e and N_p are always needed, no matter how many additional degrees of freedom the chemical space is allowed to possess.

B. Molecular grand-canonical ensemble theory

A theoretical study of chemical space must be intimately linked to a molecular GCE theory since only the latter offers the physical framework to rigorously relate molecular fluctuations in phase space to compositional fluctuations in particle space.

Generation of a conventional GCE distribution using classical statistical mechanical approaches such as grand-canonical Monte Carlo⁸ is straightforward and textbook knowledge. In addition, arbitrary λ paths connecting different molecules or functional groups via chemical or “alchemical” transformations can be used for the purpose of computing relative free energies within classical Monte Carlo or molecular dynamics (MD) calculations via thermodynamic integration or other λ -sampling techniques.^{9–16}

As noted above, a general description of chemical space involving the breaking and forming of chemical bonds is possible within a first-principles or *ab initio* theory. An *electronic* GCE theory was established decades ago within the development of Kohn-Sham (KS) density functional theory (DFT).^{17–19} Parr and Yang^{20,21} and Geerlings *et al.*²² provided a chemical interpretation of this formalism, leading to the development of conceptual DFT. Hartree-Fock theory was only recently extended to treat electronically open systems variationally.²³ Despite the grand-canonical character of electronic DFT, current applications are mostly carried out within the canonical ensemble.²⁴ Concerning fractional numbers of electrons, a controversy emerged after the seminal papers of Janak,²⁵ and of Levy and co-workers^{26,27} (see Refs. 28–36). The differentiability of electron density functionals with respect to the number of electrons has also been discussed for vanishing derivative discontinuities when the molecule is coupled to a reservoir or for very high temperatures.^{37,38} Using mixed electronic states, however, Sprik and co-workers have already demonstrated the applicability of DFT for electronic grand-canonical processes, i.e., redox processes, in the condensed phase and at finite temperature.^{39,40} Here, we seek to avoid this aspect of DFT altogether by referring to a quote from Kohn, Becke, and Parr: “For moving from one N_e to another no error is introduced if we arrange to connect all the correct integral N_e functionals.... In this way, we can imagine finite-difference formulas for all derivatives and functional derivatives that enter the formal theory in which N_e is treated as a continuous variable.”³⁴

For an entirely general molecular GCE, electrons and

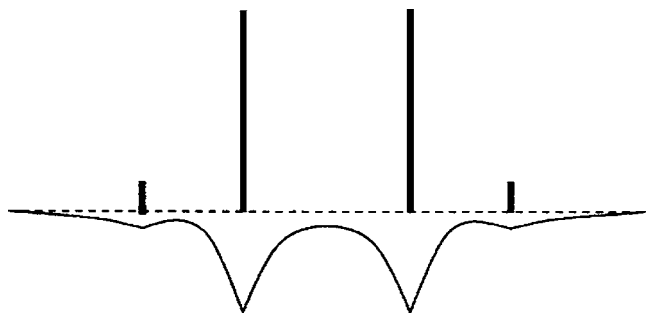


FIG. 1. Qualitative sketch of the nuclear (or proton) and electron real space distributions along the molecular axis of acetylene (C_2H_2).

nuclei must be permitted to fluctuate so that one compound can change into another. Over the last few decades, several proposals for a multicomponent (MC) non-Born-Oppenheimer DFT, wherein electrons and nuclei are treated on the same footing, have been presented.^{41–44} In particular, Capitani, Nalewajski, and Parr⁴¹ (CNP) proved a multicomponent version of the Hohenberg-Kohn theorem,¹⁷ thereby establishing the existence of an energy density functional of electronic and nuclear densities whose minimum is the ground state equilibrium geometry of the total Hamiltonian of a system of N_e electrons and N_α nuclei of type α . This theory was later revisited by Kreibich and Gross⁴³ who noted that the densities used by CNP as the basis of their MC-DFT must necessarily be constant for all isolated atoms, molecules, and solids due to the translational invariance of the Hamiltonian; they presented a MC-DFT based on densities describing the internal properties of a system. However, while these studies strive to incorporate nuclear quantum effects into the electronic DFT framework, treating electrons and nuclei on the same footing, the molecular GCE theory presented in this study is rather concerned with a grand-canonical extension of the usual Born-Oppenheimer scheme—as employed within conventional *ab initio* molecular dynamics or Monte Carlo schemes. In other words, we seek a theory that, within the accuracy of KS-DFT, describes the simultaneous variation of the electron and proton distributions in a molecule. We are not aware of any formulation of such a quantum-chemical molecular GCE theory in terms of continuous electron and proton densities.

Within the Born-Oppenheimer approximation molecules correspond to superimposed electron and nuclear (or proton) distributions, as illustrated for acetylene in Fig. 1. While N_e and N_p are independent grand-canonical variables, their canonical distributions in real space are complementary by virtue of the Hohenberg-Kohn theorems.¹⁷ CNP formulated an auxiliary variational functional Ω using electronic- and nuclear-type dependent global chemical potentials as Lagrange multipliers for the electronic and nuclear density normalization conditions. The need to consider nuclear baths for each type of nucleus, i.e., for each atomic number, can be circumvented by considering a single proton distribution which at the position of a given nucleus has to integrate to the corresponding atomic number. We show that such a theory emerges if the Lagrange multiplier corresponding to a *nuclear* chemical potential, $\mu_n(\mathbf{r})$, is treated as a local quantity, used to fix the number of protons at each point in space.

In Ref. 4, such a local nuclear chemical potential was proposed in an *ad hoc* fashion and shown to correspond to the electrostatic potential which measures, by construction, the response of a system to a positive test charge. Indeed, for inhomogeneous systems such as molecules, the notion of a local nuclear chemical potential is by no means counterintuitive. Unless the most extreme conditions prevail, an inhomogeneous system will remain inhomogeneous as the system samples its potential energy surface. Thus, for example, the work needed to insert a water molecule in the core of a globular protein will be very different from that needed to insert one somewhere in the surrounding solvent. In addition, within the context of classical density functional theory, frequently used for polymer models, local Lagrange multipliers corresponding to a local (or segmental) chemical potential are not uncommon.⁴⁵ The nuclear chemical potential at the position \mathbf{R}_I of the I th nucleus, $\mu_n(\mathbf{R}_I)$, measures the response of the system to a transmutation with respect to an *alchemical* variation of the atomic number of that atom. $\mu_n(\mathbf{R}_I)$ has therefore been termed the alchemical potential of atom I .

C. Overview of key results and interpretations of the theory

Before discussing the theory at a detailed level, we present a brief summary of the basic results of the theory. A Gibbs-Duhem analog is derived along the lines of CNP but using a continuous protonic Lagrange multiplier. As a result, we obtain a set of thermodynamic relations, in particular, the nuclear hardness, $\eta_n(\mathbf{r}, \mathbf{r}')$, and a new *molecular* Fukui function, $f_m(\mathbf{r})$, that couples electronic and nuclear degrees of freedom. The former, in complete analogy to the electronic hardness, is the second order derivative of the energy functional with respect to the nuclear density, i.e.,

$$\frac{\delta^2 E[N_e; \rho, Z]}{\delta Z(\mathbf{r}) \delta Z(\mathbf{r}')} = \frac{\delta \mu_n(\mathbf{r})}{\delta Z(\mathbf{r}')} \equiv \eta_n(\mathbf{r}, \mathbf{r}'), \quad (1)$$

implying that the nuclear hardness is a kernel. $Z(\mathbf{r})$ and $\rho(\mathbf{r})$ are the proton and electron densities, respectively. $E[N_e; \rho, Z]$ is the total molecular potential energy—a functional of ρ and Z , and a simple function of N_e .

The nuclear hardness can be expected to be of use for the study of molecular acid-base relations, or in the context of compound design, for the tuning of pK_a 's. In that sense, it might close the *protonic* conceptual gap with respect to the finite difference (deprotonated-protonated) reactivity studies on the external potential by Ayers and Parr.⁴⁶

The molecular Fukui function is the mixed second order derivative of the energy with respect to proton density and electron number, N_e ,

$$\frac{\delta^2 E[N_e; \rho, Z]}{\partial N_e \delta Z(\mathbf{r})} = \frac{\partial \mu_n(\mathbf{r})}{\partial N_e} = \frac{\delta \mu_e}{\delta Z(\mathbf{r})} \equiv f_m(\mathbf{r}). \quad (2)$$

This index can be interpreted as the response of the electronic chemical potential, μ_e , which equates the negative of electronegativity^{20,47} and, via Janak's theorem,²⁵ also the eigenvalue of the highest occupied molecular orbital (HOMO), with respect to variations of the nuclear charge distribution. By means of the Maxwell relation, however, it can likewise

be seen as the response of the electrostatic potential due to variation of the number of electrons in the system. Hence, $f_m(\mathbf{r})$ constitutes an index for changing a compound by “dialing in” a certain HOMO eigenvalue/electronegativity or electrostatic potential. For the ground state, it will be shown that the electronic hardness, η_e , is connected to the molecular Fukui functional and to the nuclear hardness via

$$\begin{aligned} \frac{\partial^2 E[N_e; \rho, Z]}{\partial N_e^2} &= \eta_e = \frac{1}{V} \int d\mathbf{r} f_m(\mathbf{r}) \\ &= \frac{1}{V^2} \int d\mathbf{r} d\mathbf{r}' \eta_m(\mathbf{r}, \mathbf{r}') \\ &= \frac{1}{V^2} \int d\mathbf{r} d\mathbf{r}' \frac{\delta^2 E[N_e; \rho, Z]}{\delta Z(\mathbf{r}) \delta Z(\mathbf{r}')}, \end{aligned} \quad (3)$$

where V is the volume.

After the derivation of the framework along the lines of CNP and the introduction of the molecular Fukui function and the nuclear hardness, we will illustrate grand-canonical applications for the case of the small, yet nontrivial many-particle system, ammonia (NH_3), using standard KS-DFT calculations: First, the two expressions for the molecular Fukui function, which are equal according to a Maxwell relation, are evaluated numerically for the geometry-optimized minimal total potential energy structure of ammonia. We find that computing $f_m(\mathbf{r})$ as the derivative of the electrostatic potential leads to practically negligible deviation from its expression as the response of the HOMO eigenvalue due to perturbing the nuclear charge distribution. This agreement numerically supports the validity of Janak’s theorem. Second, a finite temperature study is presented, illustrating how the free energy difference between neutral and oxidized ammonia at 300 K *in vacuo* can be estimated. We find that integrating over the continuous particle number path corresponding to the reversible removal of an electron leads to a value which deviates only by roughly 1 kcal/mol from what can be predicted from the harmonic vibrational spectra of NH_3 and NH_3^+ .

III. THEORY

A. From densities to Lagrange multipliers

CNP introduced in Ref. 11 a nonadiabatic variational energy functional for electronic, $\rho(\mathbf{r})$, and nuclear, $\{n_\alpha(\mathbf{r})\}$ (for every atom-type α), one particle densities and demonstrated that the external potential due to the nuclei of an optimized geometry corresponds to the grand-canonical ensemble equilibrium. Here, in contrast, we formulate a theory based on the single particle electron and proton densities, $\rho(\mathbf{r})$ and $Z(\mathbf{r})$, which, despite the translational invariance of the electronic and ionic ground-state wave function $\Psi(\{\mathbf{R}\}, \{\mathbf{r}\})$,⁴³ can be formally defined as

$$\begin{aligned} Z(\mathbf{R}) &= N_p \int d^{N_p-1} \mathbf{R} d^{N_e} \mathbf{r} |\Psi(\{\mathbf{R}_\beta\}, \{\mathbf{r}_i\})|^2 \\ &= \int d^{N_p} \mathbf{R} d^{N_e} \mathbf{r} |\Psi(\{\mathbf{R}_\beta\}, \{\mathbf{r}_i\})|^2 \sum_\beta \delta(\mathbf{R} - \mathbf{R}_\beta) \end{aligned} \quad (4)$$

and

$$\begin{aligned} \rho(\mathbf{r}) &= N_e \int d^{N_p} \mathbf{R} d^{N_e-1} \mathbf{r} |\Psi(\{\mathbf{R}_\beta\}, \{\mathbf{r}_i\})|^2 \\ &= \int d^{N_p} \mathbf{R} d^{N_e} \mathbf{r} |\Psi(\{\mathbf{R}_\beta\}, \{\mathbf{r}_i\})|^2 \sum_i \delta(\mathbf{r} - \mathbf{r}_i), \end{aligned} \quad (5)$$

where β and i enumerate all N_p protons and N_e electrons of the system to which the densities integrate, $\int d\mathbf{r} Z(\mathbf{r}) = N_p$, and $\int d\mathbf{r} \rho(\mathbf{r}) = N_e$. For simplicity, we will only consider the ionic zero-temperature limit in the present formulation. Assuming the standard molecular Hamiltonian,

$$\hat{\mathcal{H}}_{\text{total}} = \hat{\mathcal{H}}_{ee} + \hat{\mathcal{H}}_{eZ} + \hat{\mathcal{H}}_{ZZ}, \quad (6)$$

where $\hat{\mathcal{H}}_{ee}$ and $\hat{\mathcal{H}}_{ZZ}$ are the Hamiltonians of the isolated electronic and nuclear subsystems and $\hat{\mathcal{H}}_{eZ}$ is the interaction between them, the total KS-energy functional, $E[N_e; \rho, Z]$, in terms of electron and proton densities is given by

$$\begin{aligned} E[N_e; \rho, Z] &= F_{ee}[\rho] + E_{eZ}[\rho, Z] + E_{ZZ}[Z] \\ &= F_{ee}[\rho] - \int d\mathbf{r} \rho(\mathbf{r}) v(\mathbf{r}) + \frac{1}{2} \int d\mathbf{r} Z(\mathbf{r}) \bar{v}(\mathbf{r}). \end{aligned} \quad (7)$$

Here, F_{ee} is the universal Hohenberg-Kohn functional, expressible within KS-DFT in terms of the noninteracting kinetic ($T_s[\{\phi_i\}]$, where $\phi_i(\mathbf{r})$ is a KS orbital), Hartree (E_H) and exchange-correlation (E_{xc}) energies,

$$F_{ee}[\rho] = T_{ee}[\rho] + V_{ee}[\rho] \quad (8)$$

$$= T_s[\{\phi_i\}] + E_H[\rho] + E_{xc}[\rho]. \quad (9)$$

When considering a somewhat artificial “quantum-chemical” proton density, instead of densities of nuclear types, one has to take into account that protons in the same nucleus would repel each other. In order to avoid this artifact, the corresponding singularities in the nuclear Coulomb repulsion need to be excluded. It is straightforward to circumvent this problem using a formally modified external potential of the form

$$\bar{v}(\mathbf{r}) = \int d\mathbf{r}' \frac{Z(\mathbf{r}') \text{erf}[\sigma|\mathbf{r} - \mathbf{r}'|]}{|\mathbf{r} - \mathbf{r}'|}, \quad (10)$$

for the nuclear Coulomb repulsion term. Here, the parameter σ can be chosen large enough that only a vanishingly small neighborhood around $|\mathbf{r} - \mathbf{r}'| \approx 0$ is excluded.

In the spirit of CNP, one could introduce an auxiliary variational functional,

$$\Omega[N_e; \rho, Z] = E[N_e; \rho, Z] - \mu_e \left(\int d\mathbf{r} \rho(\mathbf{r}) - N_e \right) - \mu_g \left(\int d\mathbf{r} Z(\mathbf{r}) - N_p \right), \quad (11)$$

where μ_e and μ_g are the global Lagrange multipliers associated with the normalization constraints on the electronic and protonic charge densities, respectively. CNP also proved a version of the Hohenberg-Kohn theorem for such a DFT. In the thermodynamic limit, sampling these properties should yield averaged estimates for the global work required for a molecule to accept another proton or electron, respectively.

Of course, at finite temperature, and within the Born-Oppenheimer approximation, the nuclear density is given by a quantum canonical ensemble average rather than by Eq. (4),

$$Z(\mathbf{r}) = \frac{\text{Tr}[e^{-\beta H_N} \sum_I N_I \delta(\mathbf{r} - \mathbf{R}_I)]}{\text{Tr}[e^{-\beta H_N}]}, \quad (12)$$

where N_I is the number of protons in atom I , H_N is the nuclear Hamiltonian, and $\beta = 1/k_B T$. In the classical limit, a limit commonly used within standard *ab initio* molecular dynamics, this becomes a classical configurational average denoted

$$Z(\mathbf{r}) = \left\langle \sum_I N_I \delta(\mathbf{r} - \mathbf{R}_I) \right\rangle, \quad (13)$$

and, as the temperature goes to 0, the nuclear density becomes simply $Z(\mathbf{r}) = \sum_I N_I \delta(\mathbf{r} - \mathbf{R}_I)$, as expected.

Thus, at zero temperature, and when nuclei are treated at the classical level, the use of a local multiplier, $\mu_n(\mathbf{r})$, is necessary. Assuming a classical form for $Z(\mathbf{r})$, we introduce an alternative auxiliary variational functional,

$$\Omega[N_e; \rho, Z] = E[N_e; \rho, Z] - \mu_e \left(\int d\mathbf{r} \rho(\mathbf{r}) - N_e \right) - \int d\mathbf{r} \mu_n(\mathbf{r}) \left(Z(\mathbf{r}) - \sum_I N_I \delta(\mathbf{r} - \mathbf{R}_I) \right). \quad (14)$$

Here, the nuclear positions \mathbf{R}_I become variational parameters and minimization of Ω with respect to $Z(\mathbf{r})$ is tantamount to a geometry optimization—when the nuclear charges are constrained.

From Eq. (14), the Euler equations determining the exact ground-state densities $Z(\mathbf{r})$ and $\rho(\mathbf{r})$ which make Ω stationary, yield the conditions

$$\mu_e = \left. \frac{\delta E[N_e; \rho, Z]}{\delta \rho(\mathbf{r})} \right|_{Z(\mathbf{r})} = \left. \frac{\partial E[N_e; \rho, Z]}{\partial N_e} \right|_{Z(\mathbf{r})} \quad (15)$$

and

$$\mu_n(\mathbf{r}) = \left. \frac{\delta E[N_e; \rho, Z]}{\delta Z(\mathbf{r})} \right|_{N_e}. \quad (16)$$

The functional derivative of the zero-temperature electronic chemical potential μ_e is a constant for a given compound²⁰—if the electron density is forced to be the ground state density—and can be thought of as the negative value of

the electronegativity of the system.⁴⁷ Assuming that the variation in N_e concerns solely the highest occupied (HO) KS orbital with eigenvalue ϵ_{HO} , Janak's theorem,²⁵

$$\frac{\delta E}{\delta f_i} = \epsilon_i, \quad (17)$$

f_i being the occupation number of the KS-state i , implies that $\mu_e = \epsilon_{\text{HO}}$. Note that in a GCE context, where N_e varies by virtue of the fact that the electronic occupation numbers can be summed without restriction, the validity of Janak's theorem is manifestly clear. Janak's theorem is hence in agreement with the fact that $\mu_e \leq 0$ for all bound electrons. For isolated systems, the exact exchange-correlation functional exhibits derivative discontinuities at integer electron numbers²⁶ which has led to the consideration of “left” and “right-sided” derivatives of the E vs N_e curve and the use of “mean” values (averages of ionization potential and electron affinity) to approximate the derivative.²⁰ One should keep in mind that for an isolated molecule, the convex isoprotonic grand-canonical potential energy function consists of lines (for a rigid molecule) with discontinuous changes of slope at integer N_e , i.e., whenever an empty KS level becomes occupied. It should be noted, however, that while the generalized gradient approximation (GGA) (see Sec. IV A), which has been used for the applications considered herein, yields a continuous μ_e ,⁴ the self-interaction error can be expected to increase for noninteger occupation numbers.⁴⁸

According to Eqs. (7) and (15), the local nuclear chemical potential is

$$\mu_n(\mathbf{r}) = \int d\mathbf{r}' \frac{\delta F_{ee}[\rho]}{\delta \rho(\mathbf{r}')} \frac{\delta \rho(\mathbf{r}')}{\delta Z(\mathbf{r})} - \int d\mathbf{r}' \left(\frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v(\mathbf{r}') \frac{\delta \rho(\mathbf{r}')}{\delta Z(\mathbf{r})} \right) + \bar{v}(\mathbf{r}). \quad (18)$$

However, due to the Hellmann-Feynman theorem this derivative does not depend on variations in the electronic structure,⁴⁹ and can hence be identified with the (modified) electrostatic potential, $\bar{V}_{\text{ESP}}(\mathbf{r})$,

$$\mu_n(\mathbf{r}) = \int d\mathbf{r}' \frac{Z(\mathbf{r}') \text{erf}[\sigma|\mathbf{r} - \mathbf{r}'|] - \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = \bar{V}_{\text{ESP}}(\mathbf{r}). \quad (19)$$

It is well known that the derivative of the energy with respect to the atomic number corresponds to the electrostatic potential at the position of the atom,^{50–52} a fact that has been used for alternative formulations of atomic and molecular energies.^{49,53} Furthermore, the fundamental role of the electrostatic potential has been elucidated by Murray and co-workers,^{51,52} who emphasized its suitability for indexing such important properties as atomic and covalent radii, charges, and even electronegativity (which is the negative of μ_e) as the sum over all electrostatic potentials at the positions of the atoms in a molecule. In Eq. (19), we have therefore identified this important property as a grand-canonical quantity that is defined everywhere in space and that can be interpreted as the nuclear chemical potential. We call $\mu_n(\mathbf{r} = \mathbf{R}_I)$ the *alchemical* potential of nucleus I since it measures the tendency of a system to “transmute” the nucleus at \mathbf{R} .

The usefulness of this quantity for the purposes of compound design has already been illustrated within a quantum-mechanical/molecular mechanical (QM/MM) drug-design application.⁴ The relationship between the local nuclear chemical potential and the global protonic chemical potential, μ_g in Eq. (11), is elucidated in the Appendix.

Recall that when expressing the electronic chemical potential as the eigenvalue of the highest KS orbital, the approximation due to the use of a chosen exchange-correlation functional is retained. In contrast, the identification of the nuclear chemical potential as the modified electrostatic potential is exact and not limited to KS-DFT. However, the accuracy of the nuclear chemical potential corresponds to the accuracy of the electron density obtained from the given exchange-correlation functional.

B. Gibbs-Duhem

Following CNP, we now derive the Gibbs-Duhem equation for the nuclei via an auxiliary Legendre transformed energy functional $Q[\mu_e, \mu_n]$, wherein the extensive variables N_e and $Z(\mathbf{r})$ are replaced by their intensive conjugates μ_e and $\mu_n(\mathbf{r})$,

$$Q[\mu_e, \mu_n] = E[N_e; \rho, Z] - \int d\mathbf{r}(\mu_e \rho(\mathbf{r}) + \mu_n(\mathbf{r})Z(\mathbf{r})), \quad (20)$$

where we have used Eqs. (15) and (16). The total differential is then

$$dQ[\mu_e, \mu_n] = dE[N_e; \rho, Z] - N_e d\mu_e - \mu_e dN_e - \int d\mathbf{r}(Z(\mathbf{r})\delta\mu_n(\mathbf{r}) + \mu_n(\mathbf{r})\delta Z(\mathbf{r})), \quad (21)$$

where the differential of the grand-canonical total potential energy hypersurface $dE = \mu_e|_Z dN_e + \int d\mathbf{r}\mu_n(\mathbf{r})|_{N_e} \delta Z(\mathbf{r})$. Replacement of dE in Eq. (21) leads to

$$dQ[\mu_e, \mu_n] = -N_e d\mu_e - \int d\mathbf{r}Z(\mathbf{r})\delta\mu_n(\mathbf{r}), \quad (22)$$

the Gibbs-Duhem equation analog of CNP but for a continuous nuclear charge distribution. Since dQ is an exact differential, a Maxwell relation for the extensive variables and their intensive conjugates follows

$$\frac{\delta}{\delta Z(\mathbf{r})} \frac{\partial E}{\partial N_e} = \frac{\delta\mu_e}{\delta Z(\mathbf{r})} = \frac{\partial\mu_n(\mathbf{r})}{\partial N_e} = f_m(\mathbf{r}). \quad (23)$$

The function $f_m(\mathbf{r})$ measures hence the response of the electronic (nuclear) chemical potential with respect to variations in the nuclear charge density (electron number). As such, $f_m(\mathbf{r})$ is a new type of Fukui function which we call the *molecular* Fukui function, $f_m(\mathbf{r})$, as it considers both charged particle species, electrons and protons, which build up molecules (or, more generally, a complete condensed-phase system).

Now let us consider the grand-canonical total differential of the chemical potentials, $\mu_e[N_e; \rho, Z]$ and $\mu_n[N_e; \rho, Z, \mathbf{r}]$,

$$d\mu_e = \eta_e dN_e + \int d\mathbf{r}f_m(\mathbf{r})\delta Z(\mathbf{r}), \quad (24)$$

$$\delta\mu_n(\mathbf{r}) = f_m(\mathbf{r})dN_e + \int d\mathbf{r}'\eta_n(\mathbf{r}, \mathbf{r}')\delta Z(\mathbf{r}'), \quad (25)$$

where $\eta_e = \partial\mu_e/\partial N_e$ is the electronic hardness [which is inversely proportional to the global softness, which, in turn, is a polarizability index within conceptual DFT (Ref. 22)], and where we call the kernel $\eta_n(\mathbf{r}, \mathbf{r}') = \delta^2 E/(\delta Z(\mathbf{r})\delta Z(\mathbf{r}')) = \delta\mu_n(\mathbf{r})/\delta Z(\mathbf{r}')$ the *nuclear hardness*.

Introducing Eqs. (24) and (25) into the Gibbs-Duhem expression, Eq. (22), and rearranging, we find

$$dQ[\mu_e, \mu_n] = -\eta_e N_e dN_e - \int d\mathbf{r}f_m(\mathbf{r})(N_e \delta Z(\mathbf{r}) + Z(\mathbf{r})dN_e) - \int d\mathbf{r}d\mathbf{r}'\eta_n(\mathbf{r}, \mathbf{r}')Z(\mathbf{r})\delta Z(\mathbf{r}'), \quad (26)$$

which is an expansion of the differential in terms of second order energy derivatives. From Eq. (26), it becomes clear how $f_m(\mathbf{r})$ couples the nuclear with the electronic grand-canonical degrees of freedom. Regrouping with respect to particle variations gives

$$dQ[\mu_e, \mu_n] = -\left(\eta_e N_e + \int d\mathbf{r}f_m(\mathbf{r})Z(\mathbf{r})\right)dN_e - \int d\mathbf{r}\left(f_m(\mathbf{r})N_e + \int d\mathbf{r}'\eta_n(\mathbf{r}, \mathbf{r}')Z(\mathbf{r}')\right)\delta Z(\mathbf{r}), \quad (27)$$

which can be used for deriving further Maxwell relations between the third order derivatives of the energy, i.e., derivatives of η_e , $\eta_n(\mathbf{r}, \mathbf{r}')$, and $f_m(\mathbf{r})$ with respect to N_e and $Z(\mathbf{r})$.

At equilibrium, the chemical potentials must be stationary, i.e., $\delta\mu_n = d\mu_e = 0$. From Eqs. (24) and (25), it follows then that

$$\eta_e = - \int d\mathbf{r}f_m(\mathbf{r}) \frac{\partial Z(\mathbf{r})}{\partial N_e} \quad (28)$$

and that

$$f_m(\mathbf{r}) = - \int d\mathbf{r}'\eta_n(\mathbf{r}, \mathbf{r}') \frac{\partial Z(\mathbf{r}')}{\partial N_e}, \quad (29)$$

which, when incorporated in the Gibbs-Duhem analog, Eq. (27), yields the stationarity of Q ; and—if $dN_e/V = \delta Z(\mathbf{r})$ (V is the volume)—allows the electronic hardness to be related with $f_m(\mathbf{r})$ and the nuclear hardness,

$$\eta_e = - \frac{1}{V} \int d\mathbf{r}f_m(\mathbf{r}) = \frac{1}{V^2} \int d\mathbf{r}d\mathbf{r}'\eta_n(\mathbf{r}, \mathbf{r}'). \quad (30)$$

C. Second order derivatives

Using finite difference approximations to the derivatives, $f_m(\mathbf{r})$ is accessible either as the variation in the electrostatic potential, V_{ESP} , due to an infinitesimal change in N_e ; or alternatively, from Janak's theorem,²⁵ $f_m(\mathbf{r})$ is the change in

ε_{HO} due to an infinitesimal change in the nuclear charge distribution, $Z(\mathbf{r}) \mapsto Z(\mathbf{r}) + \delta Z(\mathbf{r})$. Consequently, $f_m(\mathbf{r}) \leq 0$, implying that grand-canonical variations are consistent with the extensive nature of the total potential energy functional. Hence, $f_m(\mathbf{r})$ measures how a compound's chemical potential varies upon change of its composition.

For atomic transmutations, the *alchemical* hardness $\eta_n(\mathbf{R}_I, \mathbf{R}_I)$ must be ≤ 0 in order to ensure the concavity of the isoelectronic nuclear grand-canonical potential energy surface.^{54,55} Clearly, the nuclear hardness kernel is closely related to the linear response kernel, $\delta^2 E / (\delta v(\mathbf{r}) \delta v(\mathbf{r}')) = \delta \rho(\mathbf{r}) / \delta v(\mathbf{r}')$, a well known reactivity indicator.^{22,56}

Apart from the molecular Fukui function, the electronic and nuclear Fukui functions, $\partial \rho(\mathbf{r}) / \partial N_e = \delta \mu_e / \delta v(\mathbf{r}) = f_e(\mathbf{r})$ and $\partial \mu_e / \partial \mathbf{R}_I = \partial \mathbf{F}_I / \partial N_e = f_n(\mathbf{R}_I)$, respectively, represent two well studied quantities within conceptual DFT,^{20–22} where \mathbf{F}_I is the ionic force of atom I . The electronic Fukui function, $f_e(\mathbf{r})$, can be related to $f_m(\mathbf{r})$ via

$$f_m(\mathbf{r}) = \frac{\delta \mu_e}{\delta Z(\mathbf{r})} = \int d\mathbf{r}' \frac{\delta \mu_e}{\delta v(\mathbf{r}')} \frac{\delta v(\mathbf{r}')}{\delta Z(\mathbf{r})} \quad (31)$$

$$= \int d\mathbf{r}' \frac{f_e(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (32)$$

Exploiting Maxwell-type relations, and for the sake of completeness, all combinations of second order energy derivatives involving the variables $\rho(\mathbf{r})$, $Z(\mathbf{r})$, \mathbf{R}_I , and $v(\mathbf{r})$ can be written. In addition to $f_e(\mathbf{r})$, $f_n(\mathbf{r})$, and $f_m(\mathbf{r})$, three more combinations emerge,

$$\frac{\delta^2 E}{\delta v(\mathbf{r}) \delta Z(\mathbf{r}')} = \frac{\delta \mu_n(\mathbf{r}')}{\delta v(\mathbf{r})} = \frac{\delta \rho(\mathbf{r})}{\delta Z(\mathbf{r}')}, \quad (33)$$

$$\frac{\delta^2 E}{\partial \mathbf{R}_I \delta Z(\mathbf{r}')} = \frac{\partial \mu_n(\mathbf{r}')}{\partial \mathbf{R}_I} = \frac{\delta \mathbf{F}_I}{\delta Z(\mathbf{r}')}, \quad (34)$$

$$\frac{\delta^2 E}{\partial \mathbf{R}_I \delta v(\mathbf{r})} = \frac{\delta \mathbf{F}_I}{\delta v(\mathbf{r})} = \frac{\partial \rho(\mathbf{r})}{\partial \mathbf{R}_I}. \quad (35)$$

The identity Eq. (33) is the perturbed electron density, i.e., the kernel of the electronic response to variations of the nuclear charge and is identified with the functional derivative of the nuclear chemical potential with respect to the external potential. The middle identity in Eq. (34) shows that the change in the ionic forces due to variation of the nuclear charge density is equal to the change in the nuclear chemical potential due to ionic displacement. The last relation, Eq. (35), indicates that the Fukui function analog, $\partial \rho(\mathbf{r}) / \partial \mathbf{R}_I$,⁵⁷ is equal to the change in ionic forces due to variation of the external potential. A related quantity, $\partial \phi_i(\mathbf{r}) / \partial \mathbf{R}$, has recently been proven useful as an orbital bias for scanning chemical reaction pathways within *ab initio* molecular dynamics (AIMD).^{58,59} Second order energy derivatives with respect to the same variables are well established. $\delta^2 E / \partial \mathbf{R}_I^2$ is simply the curvature of the potential energy surface, and $\delta^2 E / \delta v^2$ is the aforementioned linear response kernel.

IV. APPLICATIONS

In order to illustrate the utility of the quantities derived from the theory, electronic properties of the simple but non-trivial many-particle system NH_3 have been considered. Specifically, using finite difference, the two expressions for the molecular Fukui function have been evaluated as well as the relative free energy of oxidation of NH_3 at 300 K.

A. Technical details

All DFT calculations have been carried out with the BLYP (Refs. 60–62) functional and the plane-wave code CPMD (Ref. 63) in a box of $(8 \text{ \AA})^3$. Noninteger occupation numbers have been imposed via noninteger molecular charges—a standard feature in CPMD. Fractional protons have been included in the calculations via scaled hydrogen pseudopotentials. The Fukui function results have been obtained using the hard and analytical pseudopotentials from Ref. 64 with a large plane-wave basis-set energy cutoff of 100 Ry. For the relative free energy study numerical pseudopotentials,⁶⁵ a cutoff of 70 Ry and a massive ionic Nosé-Hoover chain thermostating scheme, in which each ionic degree of freedom is coupled to a separate thermostat chain,^{66–69} have been used. The latter maintains the nuclei at a temperature of $T=300$ K. Here, we assume that at 300 K the electronic thermal fluctuations are negligible and that the standard *ab initio* MD protocol can be applied. Heavy hydrogen atoms with mass of 7 amu have been employed allowing for relatively large time steps of 25 a.u. (≈ 0.6 fs). For every molecular configuration of the trajectories, the electronic KS-eigenvalue spectrum has been evaluated. For this study, and without any loss of consistency, the effect of a local spin density has always been neglected in order to enhance computational efficiency.

B. Molecular Fukui function

First, the variation of the nuclear chemical potential due to a change of 1% in electron number has been approximated by the finite difference in the electrostatic potential,

$$f_m(\mathbf{r}) = \frac{\partial \mu_n(\mathbf{r})}{\partial N_e} \approx \frac{V_{\text{ESP}}(N_e + dN_e, \mathbf{r}) - V_{\text{ESP}}(N_e, \mathbf{r})}{dN_e} = f_m^{V_{\text{ESP}}}(\mathbf{r}). \quad (36)$$

The result is depicted as a contour plot in Fig. 2 and predicts a maximal decrease of the HOMO eigenvalue in the vicinity of the nitrogen atom, especially at the location of the free electron pair, and becomes weaker as the distance increases.

Secondly, as displayed in the structure sketch in Fig. 2, 1% of a proton has been placed above the nitrogen along the C_{3v} -symmetry axis at distances from 0.4 to 2.4 Å in steps of 0.1 Å, keeping the relative positions of all other atoms fixed. For all of these setups and for the unperturbed ammonia Kohn-Sham eigenvalues have been computed and $f_m(\mathbf{r})$ has likewise been determined by finite difference,

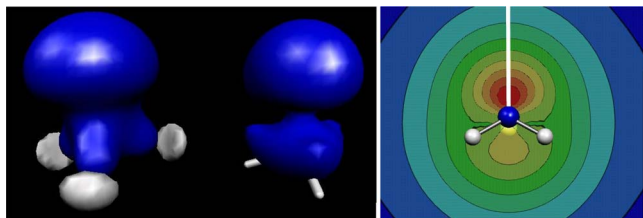


FIG. 2. (Color online) Illustration of the molecular Fukui function for ammonia. Left: isosurface of the electrostatic potential of ammonia (left) and of the molecular Fukui function (right), f_m^{VESP} , as computed from the finite difference in the electrostatic potential in Eq. (36). Right: planar contour plot of the molecular Fukui, f_m^{ESP} , function. On the C_{3v} -symmetry axis above the nitrogen perturbative positive point charges have been placed from 0.4 up to 2.4 Å distance in steps of 0.1 Å in order to compute the molecular Fukui function, f_m^e , as approximated in Eq. (37). The results for f_m^{VESP} and f_m^e are illustrated and compared in Fig. 3.

$$f_m(\mathbf{r}) = \frac{\delta \varepsilon_{HO}[Z]}{\delta Z(\mathbf{r})} \approx \frac{\varepsilon_{HO}[Z + dZ(\mathbf{r})] - \varepsilon_{HO}[Z]}{dZ(\mathbf{r})} = f_m^e(\mathbf{r}). \quad (37)$$

The resulting values of $f_m(\mathbf{r})$ are plotted in Fig. 3. The obtained unperturbed eigenvalues of ammonia are in reasonable agreement with BLYP literature values using localized basis sets.⁷⁰

While for the calculation of $f_m^e(\mathbf{r})$ a Kohn-Sham eigenvalue calculation is necessary for every value of \mathbf{r} , $f_m^{VESP}(\mathbf{r})$ can be obtained from just two wave function optimizations yielding the necessary electrostatic potentials in Eq. (36).

As expected, Fig. 3 shows that $f_m(\mathbf{r})$ is bound from below zero and that it increases in magnitude the more one approaches the high electron density region surrounding the nitrogen atom. Note that the behavior for the smallest distances is unrealistic, as these correspond to the region of the nitrogen pseudopotential. Furthermore, Fig. 3 can be seen as a numerical validation of Janak's theorem.

Given the importance of frontier orbital levels for chemical reactivity as well as for optical properties and charge transfer⁷¹ and Fermi-level properties, this result is encouraging since it indicates how to easily access $f_m(\mathbf{r})$ via Eq. (36) with KS-DFT accuracy.

C. Oxidation of ammonia

In order to further demonstrate the utility of the molecular GCE-DFT theory presented here, an AIMD calculation of

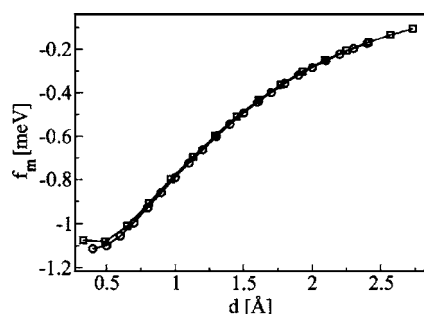


FIG. 3. Finite difference results for the molecular Fukui function. f_m is plotted as a function of the distance from the N-atom along the C_{3v} -symmetry axis in ammonia as depicted in Fig. 2. f_m^{VESP} [Eq. (36)] corresponds to squares, whereas f_m^e [Eq. (37)] corresponds to circles.

the free energy profile along the microscopic variation of particle numbers has been performed using Born-Oppenheimer MD. In principle, using successive AIMD calculations with specific combinations of the electronic and nuclear chemical potentials, relative free energies between any chemical compounds A and B , representing different points in chemical space, are accessible via thermodynamic integration,^{9,10}

$$\begin{aligned} \Delta F &= \int_B^A d\lambda \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_\lambda \quad (38) \\ &= \int_{N_e^A}^{N_e^B} dN_e \left\langle \frac{\partial E[N_e; \rho, Z]}{\partial N_e} \right\rangle_{N_e} \\ &\quad + \int d\mathbf{r} \int_{Z^A(\mathbf{r})}^{Z^B(\mathbf{r})} dZ(\mathbf{r}) \left\langle \frac{\delta E[N_e; \rho, Z]}{\delta Z(\mathbf{r})} \right\rangle_{Z(\mathbf{r})}, \quad (39) \end{aligned}$$

which for practical purposes becomes the sampled electronic chemical potential and a linear combination of sampled alchemical potentials,

$$\Delta F = \int_{N_e^A}^{N_e^B} dN_e \langle \mu_e \rangle_{N_e} + \sum_I \int_{Z^A(\mathbf{R}_I)}^{Z^B(\mathbf{R}_I)} dZ(\mathbf{R}_I) \langle \mu_n(\mathbf{R}_I) \rangle_{Z(\mathbf{R}_I)}. \quad (40)$$

In the zero-temperature limit and for $N_e^A = \{Z^A(\mathbf{R}_I)\} = 0$, this corresponds to the expressions for atomic and molecular total potential energies as presented by Wilson, Foldy, Politzer and Parr, and Berlin.^{49,50,53,72} It is shown in the Appendix how the integrand in Eq. (39) can be related to the global protonic chemical potential, μ_g in Eq. (11) via the work, $dW = \mu dN$.

However, upon reversibly transforming one compound into another via fractional amounts of particles, one is confronted with the technical question of choosing the most appropriate λ path requiring minimal computational effort for the thermal equilibration, i.e., the specific nuclei $\{I\}$ which are to be transmuted. In order to circumvent this problem for this study and without any loss of generality, we limit ourselves here to the electron particle space alone and postpone alchemical relative free energy estimates for future studies. Hence, we are exploiting the global character of the electronic chemical potential, i.e., that there is only one global grand-canonical coordinate for the removal or addition of electrons. This is different from the case of nuclear variations where, in all nontrivial cases of inhomogeneous systems such as molecules, arbitrarily many combinations and sequences of grand-canonical changes leading to the same molecular transformation can be envisioned.

We illustrate our approach for a relative free energy, trivially related to the half-reaction redox potential, corresponding to the oxidation of ammonia: $\text{NH}_3 \rightarrow \text{NH}_3^+ + e^-$. Consequently, the relative free energy expression reduces to the redox term, which (recall that μ_e is a Lagrange multiplier) via Janak's theorem²⁵ equates the sampled HOMO eigenvalue,

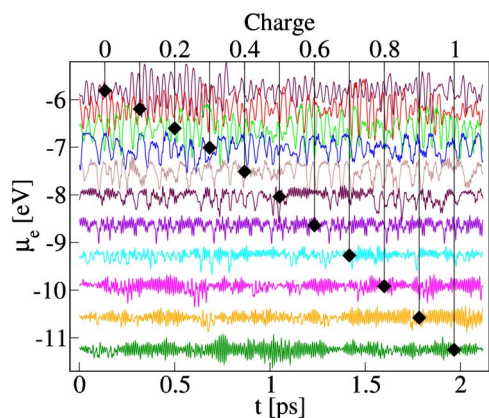


FIG. 4. (Color online) μ_e at 300 K for different charges as a function of time for 2.1 ps. The free energy profile (diamonds) for the oxidation of ammonia, i.e., $\langle \mu_e \rangle_{N_e}$ as a function of charge Q is superimposed ($Q = N_p - N_e$).

$$\Delta F_{\text{NH}_3 \rightarrow \text{NH}_3^+} = \int_{N_e=10}^{N_e=9} dN_e \left\langle \frac{\partial E}{\partial N_e} \right\rangle_{N_e} \quad (41)$$

$$= \int_{N_e=10}^{N_e=9} dN_e \langle \mu_e \rangle_{N_e} = \int_{Q=0}^{Q=1} dN_e \langle \varepsilon_{\text{HO}} \rangle_Q. \quad (42)$$

For this study, this integral is approximated by the normalized sum over the sampled HOMO eigenvalues, $\langle \varepsilon_{\text{HO}} \rangle_Q$, for the charge $Q \in \{0.0, 0.1, \dots, 0.8, 0.9, 1.0\}$ by varying N_e ($Q = N_p - N_e$). For every charge value, ammonia has been sampled at 300 K *in vacuo* for 2.1 ps. The obtained $\{\langle \varepsilon_{\text{HO}} \rangle_Q\}$ feature in Fig. 4 as a function of time, together with their time average, and represent the “grand-canonical” free energy redox profile.

The integral value for ΔF is -190.4 kcal/mol. This result is in good agreement with a free energy difference estimate obtained at the same level of theory from the (harmonic) vibrational spectra of NH_3 and NH_3^+ , respectively, $\Delta F = -189.2$ kcal/mol. The remaining deviation is attributed to the harmonic approximation and to the above mentioned shortcomings of the approximated exchange-correlation functional. As can be seen in Fig. 4, the thermal fluctuations decrease as the charge varies from 0 to 1, presumably owing to the transformation of ammonia’s double well potential energy surface into a single well for the oxidized species.

V. CONCLUSION

A multicomponent molecular grand-canonical ensemble theory incorporating the notion of a local nuclear chemical potential has been introduced as a means of sampling chemical space. We have also given a heuristic description of the latter, including a discussion of the relationship between physical and “chemical” degrees of freedom. From the theory introduced, the relations for the nuclear chemical potential, the molecular Fukui function, and the nuclear hardness have been derived. For the case of ammonia it has been illustrated how to employ grand-canonical quantities for the purpose of indexing the tuning of electronic eigenvalues or, within the context of finite temperature AIMD, for the determination of free energy differences. Sampling of chemical

space within conventional *ab initio* sampling schemes such as *ab initio* MD or *ab initio* Monte Carlo will be especially challenging since molecular grand-canonical fluctuations of both the electronic and the nuclear density require evaluation within an ergodic treatment. In analogy to the sampling of phase space, the presence of high barriers in chemical space is expected, and methods to overcome such barriers will be needed. The presented scheme is expected to be of relevance to rational compound design, relative free energy calculations from first principles, multiscale simulations where the intrinsic grand-canonical degrees of freedom can vary, and enhanced sampling methods exploiting arbitrary variable transformations.⁷³

ACKNOWLEDGMENTS

The authors are grateful to D. Andrienko for encouraging discussion during the workshop *Bridging Time and Length Scales in Materials Science and Bio-Physics*, Fall 2005, at the Institute of Pure and Applied Mathematics, UCLA, as well as to A. P. Seitsonen and C. Corminboeuf for their remarks and comments. One of the authors (O.A.v.L.) greatly acknowledges the SNF Grant No. PBEL2-110243. The other author (M.E.T.) acknowledges funding from NSF CHE-0121375 and NSF CHE-0310107.

APPENDIX: FROM THE LOCAL TO THE GLOBAL NUCLEAR CHEMICAL POTENTIAL

The global nuclear chemical potential, μ_g , measuring the likelihood to insert a proton is given via the Euler equation for the Lagrange multiplier introduced in Eq. (11) as

$$\mu_g = \frac{\partial E}{\partial N_p}. \quad (A1)$$

The associated work, or relative free energy, can be computed via $dW = \langle \mu_g \rangle dN_p$, i.e., in terms of a thermodynamic integration,

$$\Delta F = \int_{N_p^A}^{N_p^B} dN_p \langle \mu_g \rangle_{N_p}. \quad (A2)$$

The integral over all changes in $Z(\mathbf{r})$, however, must be normalized to correspond to the overall change in proton number which, for an infinitesimal but constant global change $dZ(\mathbf{r}) = dZ_g$, leads to

$$dN_p = \int d\mathbf{r} dZ(\mathbf{r}) = dZ_g \int d\mathbf{r} = V dZ_g. \quad (A3)$$

With this, the term for the local nuclear chemical potential in Eq. (39) becomes

$$\Delta F = \int d\mathbf{r} \int_{Z^A(\mathbf{r})}^{Z^B(\mathbf{r})} dZ(\mathbf{r}) \langle \mu_n(\mathbf{r}) \rangle_{Z(\mathbf{r})} \quad (A4)$$

$$= \int_{Z_g^A}^{Z_g^B} V dZ_g \int d\mathbf{r} \langle \mu_n(\mathbf{r}) \rangle_{Z(\mathbf{r})} \quad (A5)$$

$$= \int_{N_p^A}^{N_p^B} dN_p \int d\mathbf{r} \langle \mu_n(\mathbf{r}) \rangle_{Z(\mathbf{r})} \quad (\text{A6})$$

$$= \int_{N_p^A}^{N_p^B} dN_p \left\langle \int d\mathbf{r} \mu_n(\mathbf{r}) \right\rangle_{N_p}, \quad (\text{A7})$$

which must equate Eq. (A2) under the constraint that $\int d\mathbf{r} Z(\mathbf{r}) = N_p$, and consequently,

$$\mu_g = \int d\mathbf{r} \mu_n(\mathbf{r}). \quad (\text{A8})$$

$\mu_n(\mathbf{r})$ can hence be seen as an intrinsic spatial weight integrating to the global protonic chemical potential of a molecule.

- ¹C. Wolverton and A. Zunger, Phys. Rev. Lett. **81**, 606 (1998).
²A. Franceschetti and A. Zunger, Nature (London) **402**, 60 (1999).
³G. H. Jóhannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Nørskov, Phys. Rev. Lett. **88**, 255506 (2002).
⁴O. A. von Lilienfeld, R. Lins, and U. Rothlisberger, Phys. Rev. Lett. **95**, 153002 (2005).
⁵M. Wang, X. Hu, D. N. Beratan, and W. Yang, J. Am. Chem. Soc. **128**, 3228 (2006).
⁶P. Kirkpatrick and C. Ellis, Nature (London) **432**, 823 (2004).
⁷M. Dobson, Nature (London) **432**, 824 (2004).
⁸D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, New York, 2002).
⁹J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
¹⁰*Encyclopedia of Computational Chemistry*, edited by P. Ragué von Schleyer, N. Allinger, T. Clark, J. Gasteiger, P. Kollman, H.S. Schäfer III, and P. Schreiner (Wiley, New York, 1998).
¹¹B. Tidor, J. Phys. Chem. **97**, 1069 (1993).
¹²X. Kong and C. L. Brooks III, J. Chem. Phys. **105**, 2414 (1996).
¹³Z. Guo, X. Kong, and C. L. Brooks III, J. Phys. Chem. B **102**, 2032 (1998).
¹⁴K. V. Damodaran, S. Banba, and C.L. Brooks III, J. Phys. Chem. B **105**, 9316 (2001).
¹⁵R. Bitetti-Putzer, W. Yang, and M. Karplus, Chem. Phys. Lett. **377**, 633 (2003).
¹⁶J. B. Abrams, L. Rosso, and M.E. Tuckerman, J. Chem. Phys. **125**, 074115 (2006).
¹⁷P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
¹⁸M. Mermin, Phys. Rev. **137**, A1441 (1965).
¹⁹W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
²⁰R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford Science, New York, 1989).
²¹R. G. Parr and W. Yang, J. Am. Chem. Soc. **106**, 4049 (1984).
²²P. Geerlings, F. D. Proft, and W. Langenaeker, Chem. Rev. (Washington, D.C.) **103**, 1793 (2003).
²³S. Jacobi and R. Baer, J. Chem. Phys. **123**, 044112 (2005).
²⁴W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* (Wiley-VCH, Weinheim, 2002).
²⁵J. F. Janak, Phys. Rev. B **18**, 7165 (1978).
²⁶J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).
²⁷J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).
²⁸R. Vuilleumier, M. Sprik, and A. Alavi, J. Mol. Struct. **506**, 343 (2000).
²⁹M. K. Harbola, Phys. Rev. B **60**, 4545 (1999).
³⁰M. E. Casida, Phys. Rev. B **59**, 4694 (1999).
³¹L. Kleinman, Phys. Rev. B **56**, 16029 (1997).
³²J. P. Perdew and M. Levy, Phys. Rev. B **56**, 16021 (1997).
³³L. Kleinman, Phys. Rev. B **56**, 12042 (1997).
³⁴W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
³⁵A. Tachibana, Int. J. Quantum Chem. **57**, 423 (1996).
³⁶A. Tachibana and R. G. Parr, Int. J. Quantum Chem. **41**, 527 (1992).
³⁷W. Yang, Y. Zhang, and P. W. Ayers, Phys. Rev. Lett. **84**, 5172 (2000).
³⁸Y. Zhang and W. Yang, Theor. Chim. Acta **103**, 346 (2000).
³⁹I. Tavernelli, R. Vuilleumier, and M. Sprik, Phys. Rev. Lett. **88**, 213002 (2002).
⁴⁰J. Blumberger, L. Bernasconi, I. Tavernelli, R. Vuilleumier, and M. Sprik, J. Am. Chem. Soc. **126**, 3928 (2004).
⁴¹J. F. Capitani, R. F. Nalewajski, and R. G. Parr, J. Chem. Phys. **76**, 568 (1982).
⁴²N. Gidopoulos, Phys. Rev. B **57**, 2146 (1998).
⁴³T. Kreibich and E. K. U. Gross, Phys. Rev. Lett. **86**, 2984 (2001).
⁴⁴P. W. Ayers, Theor. Chim. Acta **115**, 253 (2006).
⁴⁵J. G. E. M. Fraaije, J. Chem. Phys. **99**, 9202 (1993).
⁴⁶P. W. Ayers and R. G. Parr, J. Am. Chem. Soc. **123**, 2007 (2001).
⁴⁷E. P. Gyftopoulos and G. N. Hatsopoulos, Proc. Natl. Acad. Sci. U.S.A. **68**, 3801 (1965).
⁴⁸Y. Zhang and W. Yang, J. Chem. Phys. **109**, 2604 (1998).
⁴⁹E. B. Wilson, Jr., J. Chem. Phys. **36**, 2232 (1961).
⁵⁰L. L. Foldy, Phys. Rev. **83**, 397 (1951).
⁵¹P. Politzer and J. S. Murray, Theor. Chim. Acta **108**, 134 (2002).
⁵²P. Politzer, Y. Ma, A. F. Jalbout, and J. S. Murray, Mol. Phys. **103**, 2105 (2005).
⁵³P. Politzer and R. G. Parr, J. Chem. Phys. **61**, 4258 (1974).
⁵⁴R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, J. Chem. Phys. **68**, 3801 (1978).
⁵⁵P. G. Mezey, J. Am. Chem. Soc. **107**, 3100 (1985).
⁵⁶P. W. Ayers, Theor. Chim. Acta **106**, 271 (2001).
⁵⁷R. F. Nalewajski, J. Phys. Chem. **107**, 3792 (2003).
⁵⁸L. Guidoni and U. Rothlisberger, J. Chem. Theory Comput. **1**, 554 (2005).
⁵⁹J. VandeVondele and U. Rothlisberger, J. Am. Chem. Soc. **124**, 8163 (2002).
⁶⁰A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
⁶¹R. Colle and D. Salvetti, Theor. Chim. Acta **37**, 329 (1975).
⁶²C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
⁶³J. Hutter *et al.*, Computer code CPMD, version 3.10, Copyright IBM Corp. and MPI-FKF Stuttgart 1990-2003; <http://www.cpmid.org>
⁶⁴S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
⁶⁵N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
⁶⁶N. Nosé, J. Chem. Phys. **81**, 511 (1984).
⁶⁷N. Nosé, Mol. Phys. **52**, 255 (1984).
⁶⁸W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
⁶⁹G. J. Martyna, M. L. Klein, and M. E. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).
⁷⁰U.S. Department of Commerce's Technology Administration, *National Institute of Standards and Technology*, <http://srdata.nist.gov>
⁷¹J. L. Bredas, D. Beljonne, V. Coropceanu, and J. Cornil, Chem. Rev. (Washington, D.C.) **104**, 4971 (2004).
⁷²T. Berlin, J. Chem. Phys. **19**, 208 (1951).
⁷³Z. Zhu, M. E. Tuckerman, S. O. Samuelson, and G. J. Martyna, Phys. Rev. Lett. **88**, 100201 (2002).