

## Communication: Monte Carlo calculation of the exchange energy

Roi Baer<sup>1,a)</sup> and Daniel Neuhauser<sup>2,b)</sup>

<sup>1</sup>Fritz Haber Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

<sup>2</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA

(Received 31 May 2012; accepted 25 July 2012; published online 7 August 2012)

In recent generalized Kohn-Sham (GKS) schemes for density functional theory (DFT) Hartree-Fock type exchange is important. In plane waves and grid approaches the high cost of exchange energy calculations makes these GKS considerably more expensive than Kohn-Sham DFT calculations. We develop a stochastic approach for speeding up the calculation of exchange for large systems. We show that stochastic error per particle does not grow and can even decrease with system size (at a given number of iterations). We discuss several alternative approaches and explain how these ideas can be included in the GKS framework. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4743959>]

The importance of exchange energy is deeply rooted in electronic structure theory since its early days.<sup>1–3</sup> The exchange energy can be calculated directly from an expression involving the molecular orbitals or approximately through local or semilocal density functionals.<sup>4–8</sup> Recent advances have shown that hybrid and range-separated hybrids (RSHs) are essential for applications involving charge localization-delocalization,<sup>9–11</sup> charge transfer,<sup>12</sup> and orbital gaps.<sup>13</sup> In these methods an explicit orbital exchange is added to the energy function, having the Fock-like form:

$$E_X^\gamma = -\frac{1}{4} \iint |\rho(\mathbf{r}_1, \mathbf{r}_2)|^2 u_\gamma(r_{12}) d^3 r_1 d^3 r_2, \quad (1)$$

where  $\rho(\mathbf{r}_1, \mathbf{r}_2) \equiv 2 \sum_n \phi_n(\mathbf{r}_1) \phi_n(\mathbf{r}_2)^*$  is the density matrix and  $\phi_n(\mathbf{r})$ ,  $n = 1, \dots, N/2$  are the orthonormal generalized Kohn-Sham (GKS) orbitals,  $N$  is the number of electrons, and  $u_\gamma(r_{12}) = \gamma/r_{12}$  in hybrids (usually  $\gamma \approx 0.2–0.25$ ) or  $u_\gamma(r_{12}) = \text{erf}(\gamma r_{12}/a_0)$  for RSHs, with  $\gamma \approx 0.5$ . Application of such an exchange energy functional in GKS calculations involves a considerably higher numerical cost than for KS calculations. This is especially true in calculations based on a plane-waves (PW) basis or a real space grid, where evaluation of the exchange operator is  $O(N^2)$  more expensive than the evaluation of the kinetic energy operator, as opposed to the cost of the Hartree operator which is  $O(N^{-1})$  less expensive. This scaling makes the treatment of orbital exchange in systems composed of hundreds of orbitals extremely challenging. Methods that significantly reduce the numerical effort of exchange are thus important for progress in modern DFT applications. (For references on PW/real space implementation of explicit orbital exchange, see, e.g., Refs. 14–17; further, PW implementations can use auxiliary basis techniques to solve the explicit exchange problem in a localized basis—as in, e.g., Ref. 18—leading to an  $O(N^2)$  calculation.)

Here we study the option of using Monte Carlo (MC) methods for calculating the exchange energy. The Monte

Carlo approach allows reduction of the algorithmic scaling in return for a statistical error in the calculation. The statistical error per electron is  $\delta\epsilon_X = \frac{1}{N} \sqrt{\frac{\sigma^2}{I}}$ , where  $\sigma^2$  is the variance of one Monte Carlo step and  $I$  is the number of iterations. We will give theoretical arguments and numerical evidence that the Monte Carlo process can be formulated so that  $\sigma^2$  is approximately proportional to  $N$ . The statistical error per electron  $\delta\epsilon_X$  is then proportional to  $\frac{1}{\sqrt{NI}}$ , showing that for a required level of accuracy *per electron* the number of Monte Carlo iterations  $I$  is proportional to  $\frac{1}{N}$ . Thus the larger the system, the smaller the number of Monte Carlo iterations required, at a given level of relative accuracy. This property is useful, e.g., for *ab initio* molecular dynamics which targets statistical mechanical and thermodynamical properties of liquids, water being an important example.

To start, let  $\psi_n$  be any orthonormal set of orbitals for which  $\rho(\mathbf{r}_1, \mathbf{r}_2) \equiv 2 \sum_n \psi_n(\mathbf{r}_1) \psi_n(\mathbf{r}_2)^*$ . One obviously has, in terms of the GKS orbitals:

$$\psi_n(\mathbf{r}) = \sum_{m=1}^{N/2} \phi_m(\mathbf{r}) U_{nm}, \quad (2)$$

where  $U$  is an arbitrary  $N/2 \times N/2$  unitary matrix. Our method introduces the stochastic wave function:

$$\chi(\mathbf{r}) = \sum_n e^{i\theta_n} \psi_n(\mathbf{r}), \quad (3)$$

where the phases  $\theta_n$  are  $N/2$  mutually independent real random numbers. We also use the following notations:  $\langle \rangle_I$  means an average over  $I$  random samples of phases  $\{\theta_n\}_{n=1}^{N/2}$  and  $\rightarrow$  means convergence in the limit when  $I \rightarrow \infty$ . It is straightforward to see that

$$\begin{aligned} \langle \chi(\mathbf{r}_1) \chi(\mathbf{r}_2)^* \rangle_I &\rightarrow \frac{1}{2} \rho(\mathbf{r}_1, \mathbf{r}_2) \\ \langle |\chi(\mathbf{r})|^2 \rangle_I &\rightarrow \frac{1}{2} n(\mathbf{r}), \end{aligned} \quad (4)$$

<sup>a)</sup>E-mail: roi.baer@huji.ac.il.

<sup>b)</sup>E-mail: dxn@chem.ucla.edu.

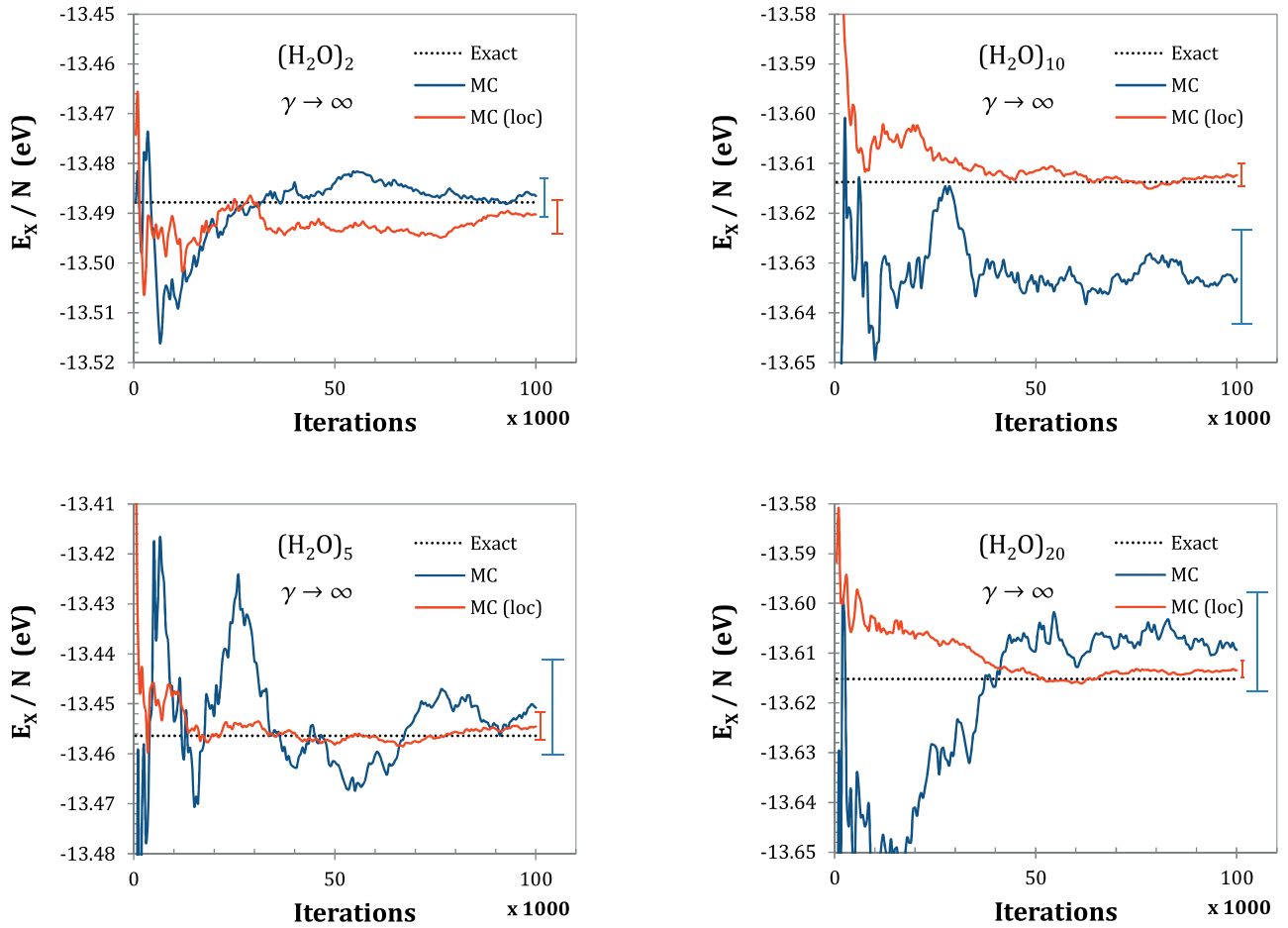


FIG. 1. The Hartree-Fock ( $u_\gamma(r_{12}) = 1/r_{12}$ ) exchange energy per electron (dotted line) of  $(\text{H}_2\text{O})_m$  ( $m = 2, 5, 10,$  and  $20$ ) and the Monte Carlo estimates based on Eq. (6) using the canonical (HF) orbitals (blue line) and localized orbitals (red line). The statistical error bars at  $10^5$  iterations are given as well (based on data of Figure 2). Note that in panels (c) and (d) the calculations with canonical orbitals are far from convergence even at  $10^5$  iterations, but we have verified that they will eventually converge if a much larger number of iterations ( $10^6$ – $10^7$ ) is used.

where  $n(\mathbf{r})$  is the electron density and

$$\begin{aligned} & \langle |\chi(\mathbf{r}_1)|^2 |\chi(\mathbf{r}_2)|^2 \rangle_I \\ & \rightarrow \frac{1}{4} n(\mathbf{r}_1) n(\mathbf{r}_2) + \frac{1}{4} |\rho(\mathbf{r}_1, \mathbf{r}_2)|^2 \\ & \quad - \sum_{n=1}^{N/2} |\psi_n(\mathbf{r}_1)|^2 |\psi_n(\mathbf{r}_2)|^2. \end{aligned} \quad (5)$$

These equations result from the statistical independence of the phases leading to  $\langle e^{i(\theta_n - \theta_m)} \rangle_I \rightarrow \delta_{nm}$  and  $\langle e^{i(\theta_n - \theta_m + \theta_k - \theta_l)} \rangle_I \rightarrow \delta_{nm} \delta_{kl} + \delta_{nl} \delta_{km} - \delta_{nl} \delta_{km} \delta_{nm} \delta_{kl}$ . As a side note, we see that  $\chi$  is not an “orbital” in the usual, single electron sense, as it represents, through ensemble averaging, the  $N$  electron system in its entirety; its norm is in general of order  $N$ . Rearranging (4) and substituting in Eq. (1) gives the following MC estimate of the exchange energy:

$$E_{X,I}^\gamma \equiv E_H^\gamma \left( \frac{1}{2} n \right) - \langle E_H^\gamma (|\chi|^2) \rangle_I - \sum_{n=1}^{N/2} E_H^\gamma (|\psi_n|^2), \quad (6)$$

where  $E_H^\gamma$  is the Hartree-like integral

$$E_H^\gamma(f) \equiv \iint f(\mathbf{r}_1) f(\mathbf{r}_2)^* u_\gamma(r_{12}) d^3 r_1 d^3 r_2. \quad (7)$$

$E_{X,I}^\gamma$  is a random variable with mean  $E_X^\gamma$  (or in the notation above  $E_{X,I}^\gamma \rightarrow E_X^\gamma$ ) and variance  $\sigma^2/I$ , where  $\sigma^2$  is estimated as

$$\sigma^2 \approx \langle E_H^\gamma (|\chi|^2)^2 \rangle_I - \langle E_H^\gamma (|\chi|^2) \rangle_I^2. \quad (8)$$

Loosely, we defined the statistical error (SE) as

$$\text{SE} = \sqrt{\frac{\sigma^2}{I}}. \quad (9)$$

In Figure 1 the Monte Carlo estimates of the exchange energy per (valence) electron are shown for clusters of water molecules of several sizes.<sup>19</sup> All calculations are done using plane waves and norm conserving pseudopotentials<sup>20</sup> and on a grid with spacing  $\Delta x = 0.5a_0$ . The single molecule calculations are enabled by using an image screening procedure.<sup>21</sup> The blue trace in each pallet of Figure 1 shows that the stochastic oscillations grow mildly as the size of the cluster grows. Thus it seems that the squared variance in the energy per electron is proportional to the number of electrons in the system,  $\sigma^2 \propto N$ . However, in the following section we show that this scaling can be reduced.

For each choice of unitary matrix  $U$  in Eq. (2)  $E_{X,I}^\gamma$  is a random variable sampled from a distribution of mean  $E_X^\gamma$

independent of  $U$  but a variance  $\sigma^2$  which does depend on it and so a good choice of  $U$  is beneficial for fast convergence. In order to understand how  $\sigma^2$  can be reduced, consider an extreme ideal case where the orbitals  $\psi_n(\mathbf{r})$  are fully localized, i.e.,  $L = 0$ , where

$$L \equiv \sum_{n \neq m} \int |\psi_n(\mathbf{r})|^2 |\psi_m(\mathbf{r})|^2 d^3r. \quad (10)$$

In this case, the variance  $\sigma^2$  is zero (i.e., one gets  $\rho(\mathbf{r}_1, \mathbf{r}_2) = \chi(\mathbf{r}_1)\chi(\mathbf{r}_2)^*$  exactly for any choice of phases) and so the exchange is obtained exactly in one iteration. Of course such an ideal situation almost never occurs in practice, but one can strive to minimize  $L$ . Alternatively, one can maximize  $L'$  where  $L = \frac{1}{4} \int n(\mathbf{r})^2 d^3r - L'$ , and

$$L'(U) = \sum_n \int |\psi_n(\mathbf{r})|^4 d^3r. \quad (11)$$

(Procedures for finding the maximally localized orbitals have been discussed extensively in the literature in other contexts.<sup>22</sup>) Indeed, in Figure 1 we show the MC samples of  $E_{X,I}^\gamma$  as a function of  $I$  when localized orthogonal orbitals are used in red. Comparing it to the samples obtained with the GKS orbitals, the statistical error is now much smaller.

We study the statistical error  $\sqrt{\sigma^2}$  per electron in water clusters of  $M$  monomers ( $(\text{H}_2\text{O})_M$ ) in Figure 2. For small clusters, localization does not make an essential difference in the statistical error; however, in the larger clusters localization has a dramatic effect. When localized orbitals are used, the statistical error per monomer for large clusters does not grow and even decreases as the system size grows. To understand this, it is instructive to examine the following idealized case, where the result can be determined analytically. Consider a system composed of  $M$  disconnected (non-overlapping) identical monomers. Clearly, one can then perfectly localize orbitals on each of the monomers separately and the MC process with these orbitals has a total variance of  $\sigma(M)^2 = M\sigma_1^2$ , where  $\sigma_1^2$  is the variance of one monomer. The statistical error per monomer is therefore  $\sigma(M)/M = \sigma_1/\sqrt{M}$ , i.e., it decreases as  $M$  grows. In connected systems one cannot expect the statistical error per electron to be reduced at this ideal rate

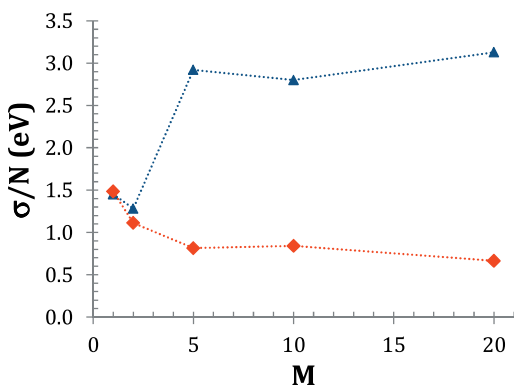


FIG. 2. Square root of the Hartree-Fock ( $u_\gamma(r_{12}) = 1/r_{12}$ ) exchange energy variance per electron for water clusters  $(\text{H}_2\text{O})_M$ ,  $M = 1, 2, 5, 10$ , and  $20$ , using HF orbitals (blue triangles) and maximally localized orbitals (red diamonds). For  $I$  MC iterations the statistical error is this variance divided by  $\sqrt{I}$  (see Eq. (9)).

of  $1/\sqrt{M}$  but if orbitals can be reasonably localized, one will see lowering of the variance per electron or per monomer. (Note that as the system grows bigger fewer Monte Carlo iterations are required for overall bulk properties, and eventually for a very large system, bulk properties, or general per-electron quantities, can be calculated by a single iteration, because then the system becomes essentially a large ensemble resulting in an ensemble average.)

We now discuss how to incorporate the MC method in GKS calculations. The GKS energy functional will include the usual deterministic terms (kinetic energy, and external, Hartree, local XC potentials) and an energy estimate obtained from an average over  $I$  iterations (Eq. (6)). For a given (constant) set of random phases one can find the orthonormal orbitals  $\psi_n$  that minimize this energy functional. This is an issue for further research, and we note here only that for the minimum search one will require the variation of the energy with respect to  $\psi_n$ , which gives, in addition to the usual deterministic terms a “stochastic” part, which for real orbitals is given by

$$\frac{\delta \langle E_H^\gamma(|\chi|^2) \rangle_I}{\delta \psi_n(\mathbf{r})} = -\text{Re} \left\langle \int \frac{|\chi(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} d^3r' \chi(\mathbf{r}) e^{-i\theta_n} \right\rangle. \quad (12)$$

Note that this quantity is obtained almost “for free” as a by-product of the calculation of  $\langle E_H^\gamma(|\chi|^2) \rangle_I$ .

Upon closing, we note two additional stochastic calculation schemes of the exchange energy. First, a version involving real phases. Define  $\eta(\mathbf{r}) = \sum_j a_j \phi_j(\mathbf{r})$  and  $\zeta(\mathbf{r}) = \sum_j b_j \phi_j(\mathbf{r})$ , with  $a_j$  and  $b_j$  real independent random variables, sampled from  $\pm 1$ , so that  $\langle a_j b_k \rangle = 0$ ,  $\langle a_j a_k \rangle = \delta_{jk}$ ,  $\langle b_j b_k \rangle = \delta_{jk}$ . The exchange energy is then  $E_X = -\frac{1}{2} \langle E_H^\gamma(\zeta(\mathbf{r})\eta(\mathbf{r})) \rangle$ , where the averaging is on the  $a$ 's and  $b$ 's. A different approach altogether is to represent the exchange energy as  $E_X = -\frac{N}{2} \langle u_\gamma(r_{12}) \rangle$ , where the average is over random 6D configurations  $(\mathbf{r}_1, \mathbf{r}_2)$  sampled from  $\rho(\mathbf{r}_1, \mathbf{r}_2)$  using a Metropolis random walk.

So far, we have described a new stochastic method for calculating the exchange energy in GKS/Hartree-Fock (HF) calculations. We have given theoretical and numerical evidence that the stochastic error in the exchange energy per electron can be made to decrease as system size grows. Future work involves two directions: first, involve the combination of the methods produced here within a GKS/HF SCF calculation. In a stochastic hybrid GKS application of exchange, only a fraction  $\gamma$  of exact exchange enters the energy equation. For a given level of statistical error, this will result in less iterations, by a factor  $\gamma^{-2}$ . The hybrid functionals have, typically  $\gamma = 0.2$  and range-separated hybrids, such as BNL (Baer-Neuhauser-Livshits, see Refs. 23 and 24) with a typical range parameter  $\gamma = 0.5$ . Both functionals have a similar reduction in the exchange energy, and therefore in the statistical error.

The second direction of future work is the choice of sampled quantity and improved sampling. We considered the variance of the total exchange energy; in practice energy differences and forces are more important, and correlated sampling is expected to be useful for retaining the same relative accuracy. The overall scaling of the suggested approach depends

on the desired quantity and the level of the localization used, and has therefore a large range. At one extreme, localization with fixed error per electron, scaling of order  $I \times N_w \log N_w$ , where  $N_w$  is the number of plane waves or grid points.  $I$  is the number of required MC samplings, a constant of order  $10^5$ – $10^6$ . Thanks to the localization, the effort in preparing a stochastic wave function is of order  $N_w$  and not  $N_w \times N$  and thus does not change the scaling. At the other extreme is a metallic system for which localization is impossible (e.g., a metallic system where the fragments are entangled over a large range), and further, a fixed total error is desired. In this case, the scaling is dominated by the preparation of the stochastic wave functions and is of order  $I \times N_w \times N$ . Other cases (e.g., localized basis set and fixed total error) lie in between these two extremes. Scaling can be improved considerably if we use *selective sampling* (so basis sets near a spatial or energy region of interest are sampled more often while most other basis functions are sampled rarely). This will be studied in future work.

The above scaling of the MC approaches can be compared with an  $N^2 \times N_w \log N_w$  scaling for non-stochastic methods not utilizing local orbitals and  $N \times N_w \log N_w$  when local orbitals can be used. Thus the stochastic method is expected to be more efficient in the limit  $N > I$ ; the crossover point which determines which method will be more or less efficient will depend on the details of implementations and on the specific system. The present simulations, of course, are quite far from the crossover point so that the Monte Carlo method took more than two orders of magnitude than the traditional non-stochastic approach.

To conclude, the method described here is the first step in constructing a fully variational GKS approach to electronic

structure based on a large given set of random phases. It opens a route for improved performance for DFT calculations involving exact exchange in very large systems.

Support by the US-Israel Binational Foundation (Grant No. 2008221) is gratefully acknowledged.

- <sup>1</sup>V. Fock, *Z. Phys.* **61**, 126 (1930).
- <sup>2</sup>J. C. Slater, *Phys. Rev.* **35**, 509 (1930).
- <sup>3</sup>W. Heitler and F. London, *Z. Phys.* **44**, 455 (1927).
- <sup>4</sup>P. A. M. Dirac, *Proc. Cambridge Philos. Soc.* **26**, 376 (1930).
- <sup>5</sup>J. C. Slater, *Phys. Rev.* **81**, 385 (1951).
- <sup>6</sup>W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- <sup>7</sup>A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- <sup>8</sup>J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- <sup>9</sup>E. Livshits and R. Baer, *J. Phys. Chem. A* **112**, 12789 (2008).
- <sup>10</sup>E. Livshits, R. Baer, and R. Kosloff, *J. Phys. Chem. A* **113**, 7521 (2009).
- <sup>11</sup>E. Livshits, R. S. Granot, and R. Baer, *J. Phys. Chem. A* **115**, 5735 (2011).
- <sup>12</sup>T. Stein, L. Kronik, and R. Baer, *J. Am. Chem. Soc.* **131**, 2818 (2009).
- <sup>13</sup>T. Stein, H. Eisenberg, L. Kronik, and R. Baer, *Phys. Rev. Lett.* **105**, 266802 (2010).
- <sup>14</sup>F. Gygi and A. Baldereschi, *Phys. Rev. B* **34**, 4405 (1986).
- <sup>15</sup>T. Todorova, A. P. Seitsonen, J. Hutter, I. F. W. Kuo, and C. J. Mundy, *J. Phys. Chem. B* **110**, 3685 (2006).
- <sup>16</sup>A. Sorouri, W. M. C. Foulkes, and N. D. M. Hine, *J. Chem. Phys.* **124**, 064105 (2006).
- <sup>17</sup>H. Sekino, Y. Maeda, T. Yanai, and R. J. Harrison, *J. Chem. Phys.* **129**, 034111 (2008).
- <sup>18</sup>M. Guidon, J. R. Hutter, and J. VandeVondele, *J. Chem. Theor. Comp.* **6**, 2348 (2010).
- <sup>19</sup>D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin, F. Calvo, J. Hernández-Rojas, and T. F. Middleton, The Cambridge Cluster Database, see <http://www-wales.ch.cam.ac.uk/CCD.html>.
- <sup>20</sup>N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).
- <sup>21</sup>G. J. Martyna and M. E. Tuckerman, *J. Chem. Phys.* **110**, 2810 (1999).
- <sup>22</sup>C. Edmiston and K. Ruedenberg, *Rev. Mod. Phys.* **35**, 457 (1963).
- <sup>23</sup>R. Baer and D. Neuhauser, *Phys. Rev. Lett.* **94**, 043002 (2005).
- <sup>24</sup>R. Baer, E. Livshits, and D. Neuhauser, *Chem. Phys.* **329**, 266 (2006).