

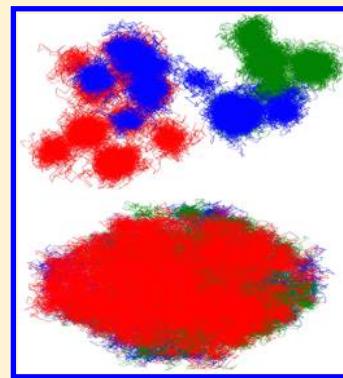
# Metropolis Evaluation of the Hartree–Fock Exchange Energy

Yael Cytter,<sup>†</sup> Daniel Neuhauser,<sup>\*,‡</sup> and Roi Baer<sup>\*,†</sup>

<sup>†</sup>Fritz Haber Center, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

<sup>‡</sup>Department of Chemistry, University of California at Los Angeles, Los Angeles, California 90095, United States

**ABSTRACT:** We examine the possibility of using a Metropolis algorithm for computing the exchange energy in a large molecular system. Following ideas set forth in a recent publication (Baer, Neuhauser, and Rabani, *Phys. Rev. Lett.* **111**, 106402 (2013)) we focus on obtaining the exchange energy per particle (ExPE, as opposed to the total exchange energy) to a predefined statistical error and on determining the numerical scaling of the calculation achieving this. For this we assume that the occupied molecular orbitals (MOs) are known and given in terms of a standard Gaussian atomic basis set. The Metropolis random walk produces a sequence of pairs of three-dimensional points ( $\mathbf{x}, \mathbf{x}'$ ), which are distributed in proportion to  $\rho(\mathbf{x}, \mathbf{x}')^2$ , where  $\rho(\mathbf{x}, \mathbf{x}')$  is the density matrix. The exchange energy per particle is then simply the average of the Coulomb repulsion energy  $v_C(|\mathbf{x} - \mathbf{x}'|)$  over these pairs. To reduce the statistical error we separate the exchange energy into a short-range term that can be calculated deterministically in a linear scaling fashion and a long-range term that is treated by the Metropolis method. We demonstrate the method on water clusters and silicon nanocrystals showing the magnitude of the ExPE standard deviation is independent of system size. In the water clusters a longer random walk was necessary to obtain full ergodicity as Metropolis walkers tended to get stuck for a while in localized regions. We developed a diagnostic tool that can alert a user when such a situation occurs. The calculation effort scales linearly with system size if one uses an atom screening procedure that can be made numerically exact. In systems where the MOs can be localized efficiently the ExPE can even be computed with “sublinear scaling” as the MOs themselves can be screened.



## INTRODUCTION

The electronic exchange energy is a basic term in density functional and Hartree–Fock theories (DFT and HFT, respectively), the first quantum correction to the classical density-based electrostatic (Hartree) energy.<sup>1–6</sup> In the local and semilocal approximations for DFT the exchange energy functional is a spatial integral over a function of the electronic density and its derivatives, which can be evaluated efficiently. Such a local/semilocal approach to DFT is useful for some applications, but extensive assessments have shown that the more elaborate nonlocal exchange is essential for quantitative description of many molecular properties, including atomization energies, bond lengths, vibrational frequencies, reaction barriers, quasiparticle energies and gaps, polarizabilities, etc.<sup>3,5,7–16</sup> Thus, it has become standard practice to incorporate nonlocal exchange into DFT calculations, mainly as a part of hybrid<sup>3</sup> or range-separated hybrid functionals.<sup>5,17–19</sup> The nonlocal exchange involves calculating a nonlocal six-dimensional (6D) integral:

$$E_X[\rho] = -\frac{1}{2} \int \int |\rho(\mathbf{x}, \mathbf{x}')|^2 v_C(|\mathbf{x} - \mathbf{x}'|) d^3x d^3x' \quad (1)$$

where  $v_C(r) = e^2/4\pi\epsilon_0 r$  is the electron–electron Coulomb repulsion potential energy and where the density matrix (DM) is a sum over occupied Kohn–Sham (KS) eigenstates:

$$\rho(\mathbf{x}, \mathbf{x}') = \sum_{n \in occ} \psi_n(\mathbf{x}) \psi_n^*(\mathbf{x}'). \quad (2)$$

Estimation of nonlocal exchange is pricy and forms the rate-determining step in many types of DFT calculations for large systems (many electrons) and large basis sets. In plane waves or grid calculations the cost of exchange energy calculation is  $O(N^2 M \log M)$ , where  $N$  is the number of electrons and  $M$  is the number of grid points, that is, near-cubic scaling. In Gaussian-type orbital basis-set calculations exchange evaluation formally has quartic  $O(M^4)$  scaling, where  $M$  is the basis set size. This, however, can be easily reduced to quadratic scaling for large systems due to the locality of the basis set functions. Further reduction to linear scaling can be achieved once  $D_{sys} > L_\rho$ , where  $L_\rho$  is the DM range, that is, the distance  $|\mathbf{x} - \mathbf{x}'|$  for which  $\rho(\mathbf{x}, \mathbf{x}')$  has decayed to negligible values, and  $D_{sys}$  is the diameter of the molecular system.<sup>20,21</sup> In most bulky systems, especially semiconductors,  $L_\rho$  is large,<sup>27</sup> and linear scaling exchange energy calculations cannot be achieved on present day computers, limiting the usefulness of linear scaling exchange calculations to low-dimensional systems.<sup>24</sup> Achieving lower scaling exchange calculation must involve new ideas that go beyond exploitation of the elusive DM sparsity.<sup>26</sup>

One such direction is to decompose the exchange integral into a sum of long (L) and short (S) range exchange terms:

$$E_X = E_X^S + E_X^L \quad (3)$$

where

Received: May 24, 2014

$$E_X^k = -\frac{1}{2} \int \int \rho(\mathbf{x}, \mathbf{x}') l^2 v^k(|\mathbf{x} - \mathbf{x}'|) d^3x d^3x', k = S, L \quad (4)$$

The short-range interaction  $v^S(r) = \text{erfc}(\gamma r)v_C(r)$  ensures that only short distances  $|\mathbf{x} - \mathbf{x}'| < n\gamma^{-1}$ , where  $\gamma^{-1}$  is the range parameter and  $n$  is a small integer, contribute to the  $E_X^S$  integral. Thus, it can be evaluated in linear scaling complexity once  $D_{\text{sys}}$  exceeds  $\gamma^{-1}$ , which can be chosen to be on the order of a few Bohrs.<sup>22</sup> By this procedure the high computational complexity of the exchange energy evaluation is deferred to the  $E_X^L$  integral over  $v^L(r) = \text{erf}(\gamma r)v_C(r)$  and as before, linear scaling “kicks in” only when  $D_{\text{sys}} > L_\rho$ , which is prohibitive.

To overcome the high complexity for evaluating  $E_X^L$  we resort to stochastic methods. These have recently been used to lower computational complexity in various types of electronic structure calculations.<sup>23,31</sup> In particular, a stochastic density functional theory (sDFT) was developed and shown capable of achieving linear scaling KS-DFT calculations.<sup>31,32</sup> In these works, a successful basic concept was developed and shown essential, namely, that the standard deviation in the *energy per particle* should be controlled instead of the total energy. It was shown that using such an approach in sDFT, which includes a converged self-consistent-field (SCF) cycle, allows calculation, to useful accuracy, of Kohn–Sham band gaps and density of states even when large energy fluctuations are present in the total energy.<sup>31</sup> Furthermore, the SCF sDFT Hamiltonian was used successfully as a basis for a stochastic approach to linear scaling GW approximations for the Dyson equation<sup>33</sup> enabling estimation of accurate quasiparticle energies. Other developments along these lines include low scaling estimation of the random phase<sup>34</sup> approximation to the correlation energy and stochastic approaches to the calculation of multiexciton generation rates in quantum dots.<sup>29,30</sup>

The achievements of sDFT are currently limited to local/semilocal Kohn–Sham functionals, and functionals that contain exact exchange make the calculation considerably more expensive. This is the main incentive for the present work (as well as an earlier attempt using a different approach<sup>28</sup>) developed in this paper where we examine a Metropolis stochastic method to calculate the exchange energy. We focus on applicability to large systems and determine the scaling of the numerical effort with system size, issues of ergodicity, and ways to accelerate the calculation.

## METHODS

As explained in the introduction, our basic premise is that in our approaches it is usefully sufficient to control the error in the exchange energy per electron (ExPE)  $e_X = (E_X)/(N_e)$ , where  $E_X$  is given in eq 1. For this, we expressed the ExPE as a sum of long and short-range terms:

$$e_X = e_X^S + e_X^L \quad (5)$$

where  $e_X^S = E_X^S/N_e$  can be evaluated using a deterministic linear scaling approach and  $e_X^L = E_X^L/N_e$  is evaluated as the expectation value of the following *random variable*:

$$e_X^L = \frac{1}{I} \sum_{m=1}^I v^L(|\mathbf{x}_m - \mathbf{x}'_m|) \quad (6)$$

where each of the  $I$  pairs of three-dimensional (3D) points  $(\mathbf{x}_m, \mathbf{x}'_m)$  is a 6D random variable distributed according to the normalized weight  $w(\mathbf{x}, \mathbf{x}') = (1/N_e)|\rho(\mathbf{x}_m, \mathbf{x}'_m)|^2$ . (The sampling

by the Metropolis method delivers directly the integral *ratio*  $(\int \rho(\mathbf{x}, \mathbf{x}')^2 v^L(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}') / (\int \rho(\mathbf{x}, \mathbf{x}')^2 d\mathbf{x} d\mathbf{x}')$  and not the integral  $\int \rho(\mathbf{x}, \mathbf{x}')^2 v^L(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}'$  itself. Since by definition  $\int \rho(\mathbf{x}, \mathbf{x}')^2 d\mathbf{x} d\mathbf{x}' = N_e$ , only the exchange energy *per particle* is accessible by the method.) Repeated sampling of the random variable in eq 6 and averaging produces an estimate for the ExPE with standard deviation given by  $\sigma$ , which can be determined by repeating the calculation several times (see section below). A useful quantity can then be defined, which we call  $\sigma_0$ , the *core standard deviation*:

$$\sigma_0 \equiv \sqrt{I}\sigma \quad (7)$$

which, because of the central limit theorem, is independent of the number of random samplings  $I$  and thus characteristic only of the electronic structure of the studied system.

An alternative approach to eq 6 would be to make a walk according to the weight  $w(\mathbf{x}, \mathbf{x}') = \text{erf}(\gamma |\mathbf{x}_m - \mathbf{x}'_m|) \rho(\mathbf{x}_m, \mathbf{x}'_m)^2$  and compute the ratio of the two averages:

$$e_X^L = \frac{\frac{1}{I} \sum_{m=1}^I v_C(|\mathbf{x}_m - \mathbf{x}'_m|)}{\frac{1}{I} \sum_{m=1}^I \frac{1}{\text{erf}(\gamma |\mathbf{x}_m - \mathbf{x}'_m|)}} \quad (8)$$

In both eq 6 and eq 8 we obtained very similar results of average and the standard deviation. Thus, we report henceforth only the method and results based on eq 6.

We use the Metropolis algorithm<sup>35</sup> to sample these pairs  $(\mathbf{x}, \mathbf{x}')$  from a random walk performed by a *random walker* in 6D space. In each step the walker attempts to assume a new position  $(\mathbf{x}, \mathbf{x})_{\text{new}}$  sampled uniformly from a 6D sphere of radius  $\Delta q$  centered on the current position  $(\mathbf{x}, \mathbf{x})_{\text{curr}}$ . The new position is accepted with probability:

$$A(\text{curr} \rightarrow \text{new}) = \min \left[ 1, \left| \frac{w((\mathbf{x}, \mathbf{x})_{\text{new}})}{w((\mathbf{x}, \mathbf{x})_{\text{curr}})} \right|^2 \right] \quad (9)$$

If the new position is rejected the walker stays in its current position. The step-size parameter  $\Delta q$  is determined by ensuring that the rate of acceptance of new positions is  $\sim 0.4$ . The tuning of  $\Delta q$  is done during a 4000 step “warm-up stage,” which also determines the initial position of the walker.

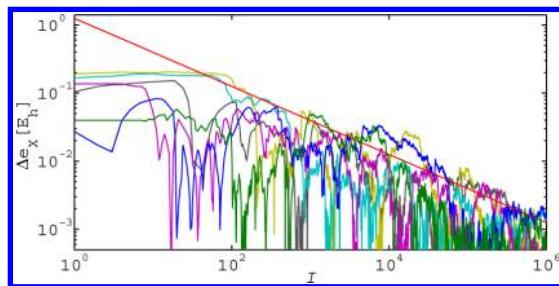
The evaluation of  $\rho(\mathbf{x}, \mathbf{x}')$  in eq 2 is based on the availability of the molecular orbitals (MOs) from the HF or DFT calculation:

$$\psi_n(\mathbf{x}) = \sum_{\mu=1}^M \phi_{\mu}(\mathbf{x}) C_{\mu n} \quad (10)$$

where  $C_{\mu n}$  are the MO coefficients and  $\phi_{\mu}(\mathbf{x})$ ,  $\mu = 1, \dots, M$  are the contracted Gaussian basis functions (CGFs) centered and localized around atom  $A_{\mu}$ . To expedite the calculation of  $\psi_n(\mathbf{x})$ , one sums only the indices  $\mu$  that refer to basis functions localized on atoms close to the point  $\mathbf{x}$ . The implications of this procedure will be studied in the next section.

For the purpose of this study, we take the MO coefficients  $C_{\mu n}$  from a conventional converged self-consistent field (SCF) Hartree–Fock calculation, performed with the Q-CHEM electronic structure package.<sup>36</sup> We chose the SBKJC pseudopotential basis set<sup>37</sup> because it enables us to avoid explicit treatment of the core electrons, considerably reducing the statistical errors. The statistical estimates are compared to the deterministic value of the exchange energy results reported in the SCF run.

The statistical estimate of the exchange energy per particle exhibits fluctuations that give rise to a standard deviation depending on the number of steps  $I$  as in eq 7 shown in Figure 1 for a silicon nanocrystal. We see that the Gaussian behavior



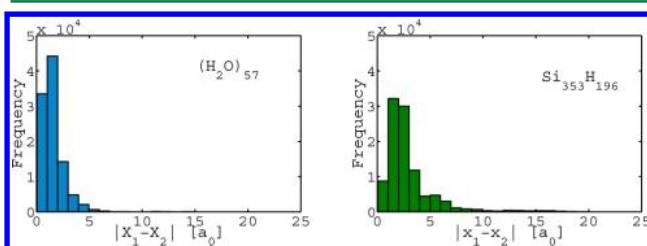
**Figure 1.** Deviance of the statistical estimate of the ExPE  $\langle e_X^L \rangle_I$  (for  $\gamma^{-1} = 0$ ) from the exact deterministic value in  $Si_{353}H_{196}$  as a function of the number of metropolis steps  $I$ . Each line represents an independent Metropolis process, and the red straight line represents eq 7, where  $\sigma_0 = 1.25E_h$  is a guiding line showing the theoretical form of the statistical convergence.

sets in only when  $I > I_0$ , where  $I_0$  is large if the distribution of single Metropolis steps is far from Gaussian as may happen when walkers fail to sample the entire space due to poor connectivity (ergodicity). One way to mitigate the effect of the initial position of the walker is to start the averaging of  $e_X$  only after a “thermalization” phase of length  $I_0$ . In the calculations below we chose  $I_0 = 4000$ .

## RESULTS

The stochastic method is demonstrated on water clusters and on hydrogen-terminated silicon nanocrystals. Two of the clusters, namely,  $(H_2O)_{31}$  and  $(H_2O)_{57}$ , were taken from the Supporting Information of ref 38. The other clusters were prepared using molecular dynamics (MD) in the following way: The water–water force field was the flexible-SPC force field with a smooth cutoff that reproduced bulk properties of water.<sup>39</sup> The MD propagation proceeded using a fourth-order

symplectic integrator with a time step of 1 fs. During the first 100 ps, frequent velocity rescaling was used to drive the system to equilibrium at the temperature of 300 K. Subsequently, a microcanonical MD trajectory was run for 200 ps, and the last time step configuration was taken as the cluster geometry we actually used.) In Table 1 we give various parameters concerning these clusters, in particular the value of the long-range ExPE as a function of the range parameter  $\gamma^{-1}$  ( $\gamma^{-1} = 0$  corresponds to the case where none of the exchange energy is short-range so all of it is long-range). The two classes of systems have very different electronic properties related to the range of quantum mechanical coherence as determined by the DM  $\rho(x, x')$ . One can get a glimpse of this range by recording the distance distribution between  $x$  and  $x'$  recorded during the metropolis walk as shown in Figure 2 for the largest silicon and



**Figure 2.** Histograms of walker distance  $|x_1 - x_2|$  for the  $(H_2O)_{57}$  water cluster (left) and the  $Si_{353}H_{196}$  silicon NC. Both systems have a diameter of approximately 1.5 nm.

water systems studied. The range of the DM for the water cluster is rather easy to characterize as  $\sim 5a_0$  because of the fast decay, but for silicon the DM tail decays slowly; significant revivals are seen even at  $15a_0$ . Another way to see that the DM for water decays fast relative to that for silicon is to observe that in Table 1 the long-range ExPE decays with  $\gamma^{-1}$  much faster for water than for silicon. As shown below the locality of the DM has important consequences on the statistical fluctuations in the exchange energy estimates.

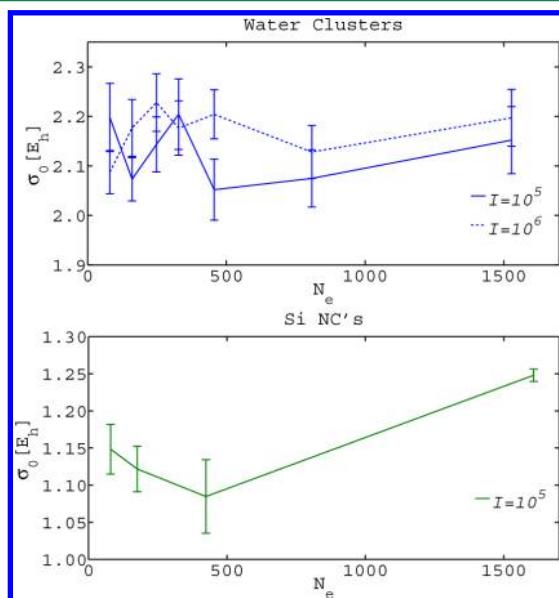
**Table 1.** Data for the  $(H_2O)_N$  Water Clusters and the Hydrogen-Capped Silicon NCs Used in the Calculations

Water clusters							
$N_{H_2O}$	$N_e^b$	$M^c$	$\gamma^{-1} = 10a_0$	$e_X^L [E_h]^a$			
				$3.3a_0$	$2.0a_0$	$0.0a_0$	
10	80	120	-0.0559	-0.1564	-0.2351	-0.491	
20	160	240	-0.0558	-0.1563	-0.2350	-0.491	
31	248	372	-0.0558	-0.1561	-0.2354	-0.488	
41	328	492	-0.0558	-0.1561	-0.2353	-0.487	
57	456	684	-0.0558	-0.1564	-0.2353	-0.492	
101	808	1212	-0.0558	-0.1560	-0.2352	-0.487	
191	1528	2292	-0.0558	-0.1561	-0.2353	-0.487	
Hydrogen-capped silicon NCs							
$N_{Si}$	$N_H$	$N_e^b$	$M^c$	$e_X^L [E_h]^a$			
				$\gamma^{-1} = 10a_0$	$3.3a_0$	$2.0a_0$	$0.0a_0$
16	16	80	160	-0.0547	-0.1392	-0.1899	-0.271
35	36	176	272	-0.0547	-0.1398	-0.1913	-0.275
87	76	424	848	-0.0547	-0.1395	-0.1908	-0.274
353	196	1608	3216	-0.0546	-0.1388	-0.1896	-0.271

<sup>a</sup>  $e_X^L$  is the deterministic value of the ExPE, given for four values of the range parameter  $\gamma^{-1}$ . <sup>b</sup>  $N_e$  is the number of electrons in the system. <sup>c</sup>  $M$  is the number of atomic orbitals in the basis.

**Sources of Error in the Metropolis Estimate of the ExPE.** There are two main sources of error in a practical Metropolis evaluation of the ExPE, namely, bias and fluctuation. Bias is a systematic error that, in the case of a properly executed Metropolis algorithm, should be zero, unless the walk is not long enough and therefore is not ergodic. We will discuss such a case below. The second source of error is the statistical fluctuations that decrease in proportion to  $1/I^{1/2}$  (see eq 7). Note that both sources of error are eliminated by increasing the number of sampling steps.

The statistical fluctuations are estimated using the standard deviation  $\sigma$  of ExPE results from 20 independent Metropolis walks of  $I$  steps. The core standard deviation  $\sigma_0$  is then calculated from  $\sigma$  and  $I$  using eq 7 and is shown in Figure 3 as a

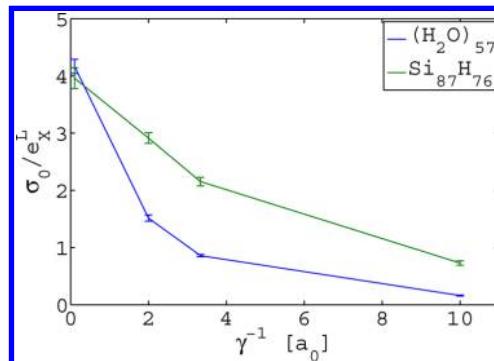


**Figure 3.** Irreducible root variance  $\sigma_0 = \sigma I^{1/2}$  in the ExPE (for  $\gamma^{-1} = 0$ ) as a function of the systems size in water clusters (top panel) and silicon NCs (bottom panel).  $\sigma_0$  was estimated by performing  $N_{\text{Met}} = 20$  independent walks of  $I = 10^5$  Metropolis steps. For water, we also show results for a random walk of  $I = 10^6$  steps.

function of system size for water and silicon clusters. It stabilizes as system size grows to a value of  $2.1E_h$  for water and  $1.2E_h$  for silicon clusters. We note that the ExPE for water is  $e_x \approx -0.49E_h$  and for silicon is  $e_x \approx -0.27E_h$ , so the relative variances, namely,  $\sigma_0/e_x$  are similar for both systems.

The variance due to the Metropolis method seems to be independent of system size for large systems (in water the variance depends on system size for small and intermediate sized clusters). This reflects the fact that the Metropolis algorithm delivers directly the ExPE and does not have the self-averaging property sometimes seen in other methods which sample the total energy and then divide by the number of particles to obtain the ExPE. Note however that this does not immediately mean that the Metropolis method is inferior to these other methods, as the CPU time is calculated in a different manner.

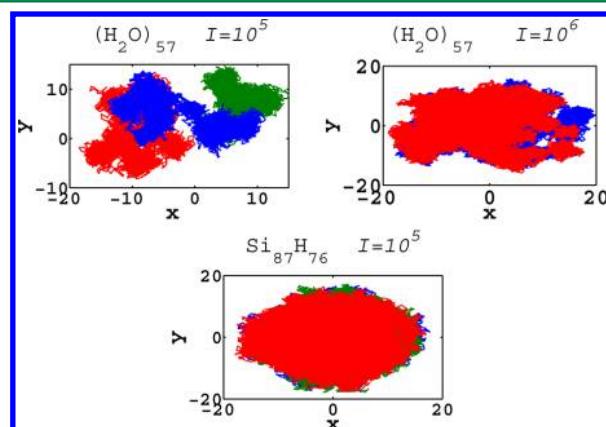
Let us now examine the effect of the range parameter  $\gamma^{-1}$  in eq 4 on the variance. Since the LR ExPE is much smaller than the full ExPE, one expects at the very least a proportional reduction in the statistical fluctuations. In Figure 4 we see that the gain is in fact larger than expected: the relative core standard deviation, that is, the ratio  $\sigma_0/e_x^L$ , decreases as a



**Figure 4.** Relative standard deviation, that is, the ratio of the core standard deviation  $\sigma_0$  to the LR ExPE  $e_x^L$  as a function of the range parameter  $\gamma^{-1}$  for a water and a silicon cluster.

function of  $\gamma^{-1}$ . We attribute this to the smoothing of the integrand, in particular the elimination of the Coulomb singularity in  $v^L(\mathbf{r})$  of eq 4. The relative core standard deviation of the water clusters benefits considerably more than that of silicon due to the lower range of the DM of the former, limiting the walker distance  $|\mathbf{x}_1 - \mathbf{x}_2|$  (see Figure 2).

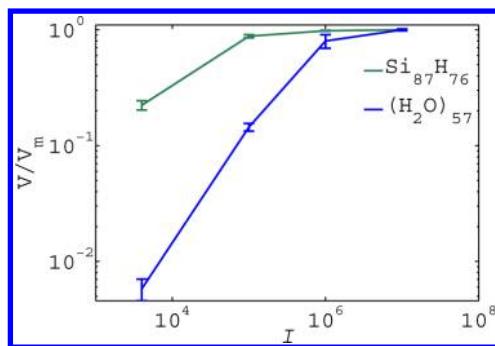
Our Metropolis method uses a *system-dependent* weight  $|\rho(\mathbf{x}, \mathbf{x}')|^2$  and a *system-independent* integrand  $v_c(|\mathbf{x} - \mathbf{x}'|)$ . Therefore, the conclusion that the relative variance for the two types of systems studied is nearly equal seems to indicate that the Metropolis sampling in both classes of systems is of a similar nature. Somewhat surprisingly, we find that this is far from true. As shown in Figure 5, a walk of  $I = 10^5$  steps in water gets



**Figure 5.** Two-dimensional projections of the  $I$  walker positions during different Metropolis walks, each represented by a different color. (upper left) Three sets of  $I = 10^5$  points for for  $(\text{H}_2\text{O})_{57}$ . (upper right) Two sets of a longer walk on the same system ( $I = 10^6$ ). (lower panel) Three sets of  $I = 10^5$  are shown for  $\text{Si}_{87}\text{H}_{76}$ .

“stuck” and does not cover the entire cluster (top left panel): it is sensitive to the initial position of a walker. In silicon (bottom middle panel) this does not happen: the walker covers to a large degree of uniformity the available space, irrespective of the initial position. This effect is due to the rather small range of the DM combined with the sparsely packed nature of the water cluster system. In silicon there is no such problem as the atoms are tightly packed and the DM has a larger range (see Figure 2). Since for water clusters, the walk of length  $I = 10^5$  is far from being ergodic, the Metropolis walk is not expected to yield the correct ExPE since it has not converged. We can measure the extent of ergodicity by estimating the ellipsoid within which the

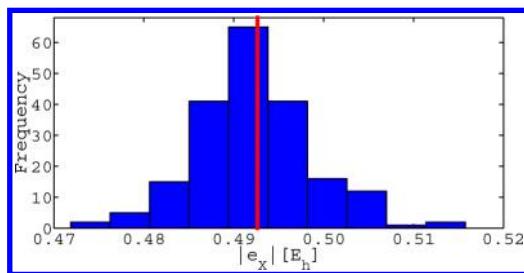
walk takes place. The center of the ellipsoid is  $\langle \mathbf{y} \rangle$ , where  $\mathbf{y} = (\mathbf{x} + \mathbf{x}')/2$  and the ellipsoid axes and corresponding widths are, respectively, the eigenvectors and eigenvalues of the  $3 \times 3$  moment of inertia tensor  $Q_{ij} = \langle y_i y_j \rangle - \langle y_i \rangle \langle y_j \rangle$ , where  $i, j = 1, 2, 3$  are two Cartesian indexes. The volume  $V = (4\pi)/(3) (\det(Q))^{1/2}$  of this ellipsoid can be compared to the converged volume  $V_m$  (estimated as the volume of the ellipsoid for a large number of walkers). A ratio  $V/V_m$  smaller than 1 indicates a problem in ergodicity. In Figure 6 we plot this ratio for a water



**Figure 6.** Volume measure of ergodicity for the random walk in a Si NC and a water cluster as a function of the length of the walk.

cluster and a silicon NC. As  $I$  increases the volume ratio approaches unity. This happens much faster in the silicon clusters than in water. Thus, unless importance sampling is used the water calculation requires a considerably higher value of  $I$ .

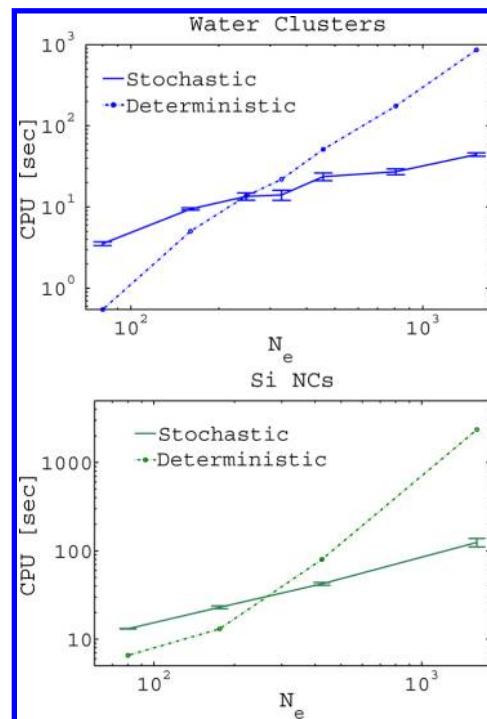
If we increase the number of steps of the walk in the water cluster we find that the trajectory does eventually cover the entire space. However, this will take a considerably longer walk, and the estimate for the ExPE by the Metropolis random walk is not reliable unless the walk is ergodic. It is interesting to note, however, despite that full ergodicity is not achieved at  $I = 10^5$  steps, the estimate of the ExPE for the water clusters is very close to the exact value, not showing a visible bias. The histogram displayed in Figure 7 shows that the short metropolis



**Figure 7.** A histogram, summarizing the results of 200 Metropolis estimates of the ExPE with  $\gamma^{-1} = 0$  for  $(\text{H}_2\text{O})_{57}$ . Each run includes 96,000 steps. The red line represent the exact value of the exchange energy per electron.

process happens to give the correct value of the integral. This happens probably because all the disparate regions have very similar exchange energy per electron. In a less homogeneous system we expect a bias and larger standard deviations. The problem of getting stuck is left as a problem for future study. A likely solution will be based on a multiscale strategy whereby the system is divided to many regions, in each of which a short trajectory is used; the choice of which region to sample will be based on the rates of going from one region to another.

**Computational Cost.** The CPU time  $T$  for calculating one Metropolis iteration is determined primarily by the CPU time needed to estimate the DM at a given pair of 3D points  $(\mathbf{x}, \mathbf{x}')$  and this in turn is determined by the CPU time to compute the value of each of the  $N_e$  occupied MO  $\psi_n(\mathbf{x})$  at points  $\mathbf{x}$  and  $\mathbf{x}'$ . Each MO is expressed as a linear combination of  $M$  atomic orbitals (AOs) and since there are  $N_e$  such orbitals  $T$  is proportional to  $MN_e$ . This quadratic scaling is not really essential and can be reduced by exploiting the fast Gaussian decay of the basis-set functions: only AOs belonging to atoms that are closer to  $\mathbf{x}$  than a threshold distance  $D_{th}$  need be taken into account. We set  $D_{th}$  to be the largest standard deviation of an atom's CGF, times a factor  $f$  determining the tolerated error. Taking  $f = 5$  is equivalent to a tolerated error of  $10^{-16}$ , which is the double precision accuracy. In applications of this approach, we found that the average number of atoms  $N_A$  required for each MO evaluation saturates as system size grows, where asymptotically  $N_A = 40$  ( $N_A = 80$ ) for the water (silicon) systems. With this technique we achieve asymptotic linear scaling behavior  $T \propto N_e$  as seen in Figure 8 for both the water



**Figure 8.** CPU time as a function of number of electrons for water clusters (upper) and Silicon NCs (lower). The reported time is for  $I = 10^5$  Metropolis steps. Error bars indicate the uncertainty in the time (determined by measuring the scatter in different runs). The CPU time per SCF iteration for the deterministic Hartree–Fock calculation is shown on the graph as well. This time is a rough estimate for the CPU time of a deterministic exchange energy calculation.

and the silicon systems. The saturation is faster for the water clusters than for silicon, due to the larger value of  $D_{th}$  required for the silicon AOs relative to the oxygen AOs. This linearity can be compared to the time for deterministic evaluation, which seems to scale in a Q-CHEM SCF calculation as  $N_e^{2.5}$ .

## SUMMARY AND DISCUSSION

In this paper we have examined the use of a Metropolis algorithm for computing the ExPE in a large molecular system,

given the molecular orbitals expressed as linear combinations of Gaussian-type atomic basis sets. As a first step, we divide the exchange energy and correspondingly the Coulomb interaction into long- and short-range parts, determined by a range parameter  $\gamma^{-1}$ . The short-range part can in principle be computed in a linear scaling manner, and we do not discuss it further here. The long-range ExPE is estimated using Metropolis walk that samples the distribution  $\rho(\mathbf{x}, \mathbf{x}')$ <sup>2</sup> and averages the long-range Coulomb repulsion energy  $v^L(|\mathbf{x} - \mathbf{x}'|)$  (using eq 6). The properties of the method were studied using two systems of very different chemical and electronic structures, namely, large water clusters and silicon clusters.

It was seen that in silicon the Metropolis walk efficiently samples the entire available space and gives a good estimate of the ExPE with a core standard deviation  $\sigma_0$  that is asymptotically independent of system size.

Similar results to those of silicon were obtained for the water clusters; however, in these systems a problem was observed, namely, the random walk does not easily cover the entire cluster and instead tends to get stuck in disparate regions. For finite walks this impairs ergodicity, which may cause a bias in the Metropolis-based estimate of the exchange energy. To overcome this we must make a considerably longer walk to restore ergodicity. The reason behind this problem is the combination of sparse packing of the water cluster and the relatively small range of the DM. In an attempt to overcome such a bottleneck in the metropolis walk we considered changing the weight function to a more delocalized one. We tried several ideas, including those developed successfully for related problems (e.g., those of ref 25) combining the electron density (or simple approximations to it) with the Coulomb operator as a replacement for the Metropolis weight. We found that, while these procedures were effective in eliminating the ergodicity problem, they had an adverse effect on the standard deviation, which grew considerably. Thus, it seems that one should not completely abandon the squared-DM as a weight; perhaps a combination of the two approaches can be beneficial in some cases, although not in the systems we study here. We leave this issue of overcoming the slow ergodicity to future work.

The efficacy of the range separation was demonstrated by showing that the ratio  $\sigma_0/e_X^L$  decreases with increasing  $\gamma^{-1}$ . We attributed this reduction to the smoothing of the  $e_X^L$  integrand and in particular the removal of the Coulomb singularity. Thus, overall the separation of the exchange integral into long- and short-range components is very beneficial. Beyond the mere fact that the Metropolis is now sampling a smaller function, it is sampling a smoother one as well; a range parameter of  $\gamma^{-1} = 10a_0$  reduces the statistical error by 2 orders of magnitude relative to  $\gamma^{-1} = 0$ .

One conclusion of this study concerns the dependence of the fluctuation size on system size. In the stochastic method of ref 31 one computes the total energy and then divides by the number of electrons to obtain the *energy per electron*. This procedure then makes the fluctuation size proportional to the inverse square-root of the number of electrons: the variance in the energy per particle decays as system grows. This in turn allows for the “sublinear scaling” calculation of the energy per electron. In the present method, fluctuations behave differently. Because the Metropolis algorithm produces the energy per electron directly (not through the total energy) the variance of the procedure is system-size independent (as seen Figure 3). The result is that the number of iterations required for

achieving a given fluctuation in the ExPE is independent of system size, and this leads to the conclusion that the numerical complexity of the Metropolis algorithm is linear scaling with system size. Sublinear scaling could be reached if it was possible to localize the MOs since then one could compute each value of  $\rho(\mathbf{x}, \mathbf{x}')$  in sublinear scaling effort. However, this regime will be reached in much larger systems as DM sparsity is illusive.

The present paper studies the possibility of computing the exchange energy, given a DM. In application to DFT these techniques must be augmented so as to enable incorporation into a self-consistent scheme, as has been done recently for local exchange potentials (see ref 28). We leave this important and difficult development for future work.

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*Email: dxn@chem.ucla.edu.

\*Email: roi.baer@huji.ac.il.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Professor E. Rabani from Tel Aviv University for supplying us with the structures of the water and silicon clusters and for his helpful comments on the manuscript. We gratefully acknowledge support from the U.S.-Israel Binational Science Foundation. D.N. is grateful for support from the NSF.

## ■ REFERENCES

- (1) Fock, V. Z. *Phys.* **1930**, *61*, 126.
- (2) Slater, J. C. *Phys. Rev.* **1930**, *35*, 509–529.
- (3) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (4) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 8207–8215.
- (5) Ikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 3540–3544.
- (6) Livshits, E.; Baer, R. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2932–2941.
- (7) Seidl, A.; Gorling, A.; Vogl, P.; Majewski, J. A.; Levy, M. *Phys. Rev. B* **1996**, *53*, 3764–3774.
- (8) Sodupe, M.; Bertran, J.; Rodriguez-Santiago, L.; Baerends, E. J. *J. Phys. Chem. A* **1999**, *103*, 166–170.
- (9) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (10) Aulbur, W. G.; Stadele, M.; Gorling, A. *Phys. Rev. B* **2000**, *62*, 7121–7132.
- (11) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. T. *Science* **2008**, *321*, 792–794.
- (12) Baer, R.; Livshits, E.; Salzner, U. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85–109.
- (13) Barone, V.; Hod, O.; Peralta, J. E.; Scuseria, G. E. *Acc. Chem. Res.* **2011**, *44*, 269–279.
- (14) Isseroff, L. Y.; Carter, E. A. *Phys. Rev. B* **2012**, *85*, 235142.
- (15) Kronik, L.; Stein, T.; Refaelly-Abramson, S.; Baer, R. *J. Chem. Theory Comput.* **2012**, *8*, 1515–1531.
- (16) Isborn, C. M.; Mar, B. D.; Curchod, B. F.; Tavernelli, I.; Martínez, T. J. *J. Phys. Chem. B* **2013**, *117*, 12189–12201.
- (17) Savin, A. In *Recent Advances in Density Functional Methods Part I*; Chong, D. P., Ed.; World Scientific: Singapore, 1995; p 129.
- (18) Heyd, J.; Scuseria, G. E. *J. Chem. Phys.* **2004**, *120*, 7274–7280.
- (19) Baer, R.; Neuhauser, D. *Phys. Rev. Lett.* **2005**, *94*, 043002.
- (20) Schwegler, E.; Challacombe, M. *J. Chem. Phys.* **1996**, *105*, 2726–2734.
- (21) Ochsenfeld, C.; White, C. a.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663.
- (22) Izmaylov, A. F.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2006**, *125*, 104103.

- (23) Thom, A. J.; Alavi, A. *Phys. Rev. Lett.* **2007**, *99*, 143001.
- (24) Booth, G. H.; Smart, S. D.; Alavi, A. *Mol. Phys.* **2014**, *1–15*.
- (25) Willow, S. Y.; Kim, K. S.; Hirata, S. *J. Chem. Phys.* **2012**, *137*, 204122.
- (26) Willow, S. Y.; Hermes, M. R.; Kim, K. S.; Hirata, S. *J. Chem. Theory Comput.* **2013**, *9*, 4396–4402.
- (27) Ge, Q.; Gao, Y.; Baer, R.; Rabani, E.; Neuhauser, D. *J. Phys. Chem. Lett.* **2014**, *5*, 185–189.
- (28) Baer, R.; Neuhauser, D. *J. Chem. Phys.* **2012**, *137*, 051103.
- (29) Baer, R.; Rabani, E. *Nano Lett.* **2012**, *12*, 2123.
- (30) Baer, R.; Rabani, E. *J. Chem. Phys.* **2013**, *138*, 051102–4.
- (31) Baer, R.; Neuhauser, D.; Rabani, E. *Phys. Rev. Lett.* **2013**, *111*, 106402.
- (32) Neuhauser, D.; Baer, R.; Rabani, E. *J. Chem. Phys.* **2014**, *141*, 041102.
- (33) Neuhauser, D.; Gao, Y.; Arntsen, C.; Karshenas, C.; Rabani, E.; Baer, R. *Phys. Rev. Lett.* **2014**, *113*, 076402.
- (34) Neuhauser, D.; Rabani, E.; Baer, R. *J. Phys. Chem. Lett.* **2013**, *4*, 1172–1176.
- (35) Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
- (36) Shao, Y.; et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (37) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026.
- (38) Richard, R. M.; Herbert, J. M. *J. Chem. Theory Comput.* **2013**, *9*, 1408–1416.
- (39) Fennell, C. J.; Gezelter, J. D. *J. Chem. Phys.* **2006**, *124*, 234104.