LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Linear scaling first-principles molecular dynamics with plane waves accuracy

J.-L. Fattebert, F. Gygi

October 11, 2005

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Linear Scaling First-Principles Molecular Dynamics with Plane Waves accuracy

Jean-Luc Fattebert, François Gygi

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551

(Dated: September 30, 2005)

We propose a real-space finite differences approach for accurate and unbiased O(N) Density Functional Theory molecular dynamics simulations based on a localized orbitals representation of the electronic structure. The discretization error can be reduced systematically by adapting the mesh spacing, while the orbitals truncation error decreases exponentially with the radius of the localization regions. For regions large enough, energy conservation in microcanonical simulations is demonstrated for liquid water. We propose an explanation for the energy drift observed for smaller regions.

I. INTRODUCTION

Density Functional Theory (DFT) is now a well established model for numerical simulations in condensed matter and is used to study a wide range of phenomenas (see *e.g.* Ref. [1–3]). Modern supercomputers allow DFT simulations with a few hundred atoms. That number depends on the time scale of interest, as well as on the number of degrees of freedom required to describe accurately the various atomic species composing the physical system. Most application codes are limited in problem size by the total number of arithmetic operations which grows like $O(N^3)$, where N is the number of electrons. Thus the size of tractable problems — using the standard Plane Wave (PW) approach for example — cannot grow as fast as computer power. This issue has motivated research on linear scaling approaches, that is numerical algorithm with O(N) complexity. Several new ideas were published at the begining of the 1990s (for a review, see *e.g.* Ref. [4]). For accurate first-principles methods however, exploring such ideas has remained a challenging research area. Indeed linear scaling methods typically become competitive for large problems that were beyond the computer ressources available to most researchers until recently. Even now, achieving manageable timings for large systems requires a very efficient parallelization. Thus scalability on a large number of processors is also a very desirable property in order to address large problems where linear scaling algorithms becomes relevant. As far as O(N) molecular dynamics (MD) simulations are concerned, several attempts by various authors have been plagued by a systematic negative drift in total energy when introducing localization constraints on the orbitals [5–9]. This means an artificial loss in energy over time which is hard to compensate for without affecting the thermodynamics of the system.

In this Letter we demonstrate for the first time the possibility of carrying out energy conserving first-principles molecular dynamics with PW accuracy in $O(N)$ operations. Our approach is in the line of traditional PW calculations where the solution of the Kohn-Sham (KS) equations is represented by a set of occupied electronic orbitals described in a very accurate numerical basis set.

Instead of dealing with extend electronic states, we represent the same occupied space by a set of nonorthogonal orbitals that can be localized. This general approach was first proposed by Galli and Parrinello[10]. The electronic structure of a physical system is given by N nonorthogonal orbitals $\{\phi\}_{i=1}^N$ that spans the invariant subspace of the N doubly occupied electronic states (neglecting the spin) and minimize the DFT energy functional

$$
\begin{aligned}
E_{KS}[\{\phi_i\}_{i=1}^N] \quad &= \sum_{i,j=1}^N \left(S^{-1}\right)_{ij} \int_\Omega \phi_i(\mathbf{r}) \left(-\nabla^2\right) \phi_j(\mathbf{r}) d\mathbf{r} \\
&+ \frac{1}{2} \int_\Omega \int_\Omega \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 + E_{XC}[\rho] \\
&+ \sum_{i,j=1}^N 2 \left(S^{-1}\right)_{ij} \int_\Omega \phi_i(\mathbf{r})(V_{ext}\phi_j)(\mathbf{r}) d\mathbf{r}.
\end{aligned}
$$

S is the overlap matrix between orbitals and ρ is the electronic density $\rho(\mathbf{r}) = 2\sum_{i,j=1}^N (S^{-1})_{ij}\phi_i(\mathbf{r})\phi_j(\mathbf{r})$.

We use orbitals described by their value on a real-space mesh and discretize the Laplacian by finite differences (FD) as proposed in Ref.11. FD was initially proposed as a simple and efficient method for electronic structure calculations on parallel architectures.[12] A real-space discretization is also a very natural approach to describe orbitals localized in real-space and truncated beyond a cutoff radius R_c. Other similar algorithms have been proposed in the literature, based on various discretization schemes for the Laplacian and the representation of the electronic wave functions[13–15]. Truncation of the orbitals outside localization regions (LR) is justified by the exponential decay of the generalized Wannier functions for systems with a finite band gap[16] and leads naturally to O(N) scaling. One original and important feature of our algorithm is the introduction of adaptive localization centers (ALC) for the LR. This improves accuracy and allow general molecular dynamics (MD) simulations.[9]

In this Letter we demonstrate how such a linear scaling approach is a practical alternative to PW calculations for arbitrary accuracy first-principles MD for very large systems. Our linear-scaling approach allows a total control of the accuracy. First, the mesh spacing can be reduced

to ensure that the problem is solved within an acceptable discretization error. We use a compact fourth order FD approach — Mehrstellen scheme — so that the error on the energy is reduced by a factor 16 when the grid size is doubled. This is an efficient approach for massively parallel computation[17]. Second, the truncation radius used to limit the extend of the localized orbitals can be enlarged in order to ensure results within a certain tolerance compared to the standard $O(N^3)$ approach, i.e. without localization constraints. This truncation error decays exponentially with the localization radius[9].

The energy functional is discretized on a uniform real-space mesh that covers the whole computational cell. The potentials, wave functions and electronic density are represented numerically by their values at the nodes of a common uniform grid. We use periodic boundary conditions. A filtering procedure is applied to the pseudopotentials before interpolating these at the grid points in order to reduce dependance on grid position [17]. The results presented in this paper were obtained using the GGA-PBE exchange-correlation functional[18]. We use norm-conserving pseudopotentials in the Kleinman-Bylander form [19].

The gradient of the energy functional (1) determines the steepest descent direction. This direction is combined with a multigrid preconditioner which provides a convergence rate independent of the discretization mesh spacing. An even faster convergence is achieved by using the preconditioned steepest descent directions in a block version of Anderson's extrapolation scheme as described in Ref. [9]. This algorithm has the property of not involving any linear combination among the trial wave functions localized in different regions and thus is very appropriate to deal with localized orbitals.

Linear scaling is achieved by an iterative minimization of the energy functional with localization constraints on the orbitals. This is done by truncating the corrections computed at each step in order to preserve a strict localization of all the trial localized orbitals. Inverting the matrix S is done without any truncation, and thus in $O(N^3)$ operations. With an efficient parallelization this operation is negligible in all the calculations we have been carrying out so far. Note that unlike several other linear scaling algorithms, N is minimal here: it is equal to the number of occupied states.

A key element in achieving accurate electronic structure calculations at minimal cost is the adaptivity of the localization centers. This is done by iteratively repositioning these centers at the centroids of charge of each orbital.[9] This enables an accurate solution without including any unoccupied orbitals in the calculations for systems with a finite band gap. In the case of water, only 4 doubly occupied states are computed for each water molecule. For comparison, methods based on numerical atomic orbitals computed and parametrized beforehand require several empty orbitals per molecule (see e.g. Ref. [20]). On the other hand our approach requires obviously more degrees of freedom per orbital. This provides more
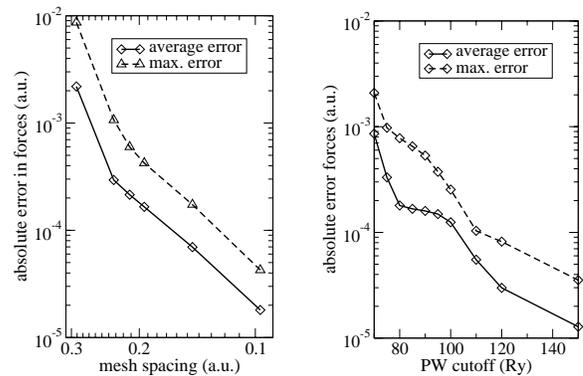

FIG. 1: Discretization error on the forces for the finite differences "Mehrstellen" and the Plane Wave approaches for $(H_2O)_{32}$ (average and maximum over sample). The reference values are the result of a fully converged PW calculation (400 Ry).

flexibility and avoids any reparametrization according to environmental factors. ALC are essential for molecular dynamics; they can adapt to the environment and move along with the dynamics. This also removes the difficulty of choosing appropriate localizations centers in general.

Since our reference approach is the PW method, we evaluate the discretization error by computing the difference between our numerical results and a fully converged PW calculation. We choose a sample of 32 water molecules in the liquid phase as test system. The measured discretization error on the energy for the FD approach shows a convergence rate of $O(h^4)$, in agreement with the theoretical convergence rate. Figure 1 shows the measured discretization error on forces as a function of mesh spacing. For comparison, we also show the results for PW.

It is desirable to define an equivalent energy cutoff for real-space FD discretizations. PW implementations usually use two different cutoffs: one for the wave functions, and a larger one for the electronic density. Ab initio molecular dynamics of water are usually carried out with a PW cutoff of 70 to 85 [Ry] for the wave functions[21, 22]. Based on the comparison of average discretization errors on forces between the two approaches (Fig.1), we establish that an 80 Ry PW cutoff is equivalent to a mesh spacing of 0.21 Bohr for the FD approach. For the 32 water molecules cell, it corresponds to a grid of 88 points in each direction, while we use grids of 56 and 110 for the wave functions and electronic density representation respectively in the PW calculation. We use these parameters in the remaining of the paper.

A previous study showed that the truncation error generally decays exponentially with the radius of the localization regions. For liquid water, an accuracy better than 10^{-3} Bohr — i.e. of the order of magnitude of the discretization error — is reached for $R_c = 8$ Bohr.[9] Solving the KS equations within the discretization accuracy is not enough however to guarantee a sufficiently good

TABLE I: Measure of energy drift and number of SC steps required for convergence.

Localization radius (Bohr)	Energy drift (mHa/atom/ps)	# SC iteration/MD step
8	-0.222 (-47 K/ps)	29
9	-0.103 (-22 K/ps)	23
10	+0.001 (0 K/ps)	14

conservation of the total energy. To demonstrate the applicability of our O(N) approach to large scale Born-Oppenheimer MD problems, we have simulated liquid water at ambient conditions for a 512 molecule sample in a cubic cell of side length 46.98 Bohr Three different localization radii were used: 8, 9 and 10 Bohr. At each ionic step, the wave functions were optimized until the energy change between two consecutive steps is smaller than $8 \cdot 10^{-6}$. We use a time step of 0.242 fs. As shown in Table I, the number of iterations required to reach this tolerance decreases with increased localization radius. The results with localization radius 8 and 9 Bohr show a negative drift in the total energy. This drift, measured by a linear fit over 0.1 ps, decreases with increased localization radius and becomes exceedingly small for 10 Bohr (see Table I). Remarkably, the number of self-consistent (SC) steps required to achieve convergence at each MD steps also drops at 10 Bohr to reach a similar value as the one obtained for an $O(N^3)$ calculation with 64 molecules (as reported in [9]).

We take a closer look at the total energy during the simulation and propose an explanation for the energy drift. For $R_c = 10$ Bohr the total energy is not perfectly constant, but fluctuates smoothly during the entire simulation. For smaller radii (8 and 9 Bohr), the smooth fluctuations are interrupted by sharp drops in energy at various points during the 0.1 ps runs. Simultaneously, the number of SC iterations required to reach the tolerance criterion sharply increases, while the average and maximum displacements of the ALC also significantly increase (see Fig.2). We interpret this phenomena as the transition between two local minima of the KS energy, falling into a local minima of slightly lower energy. This transition is followed by a *global* reorganization of the orbitals — whose signature is a much larger than average displacement for most ALC — that require more SC iterations and generates discontinuities in the ALC trajectories. This is accompanied by a drop in the KS energy which is not taken into account in the forces computation, and thus produces a net decrease in total energy. This energy drift corresponds to as much as 48 [K/ps] for $R_c = 8$ Bohr. Going towards larger localization radii reduces this effect. For $R_c = 10$ Bohr, no such discontinuity occurs over a 0.1 ps test.

Note that the increases in ALC displacements, while clearly visible, remain quite small (< 0.1 Bohr). They correspond to changes in shape of the localized orbitals that are hard to see on a contour plot. For illustration purposes, we show in Fig. 3 the contour plots of
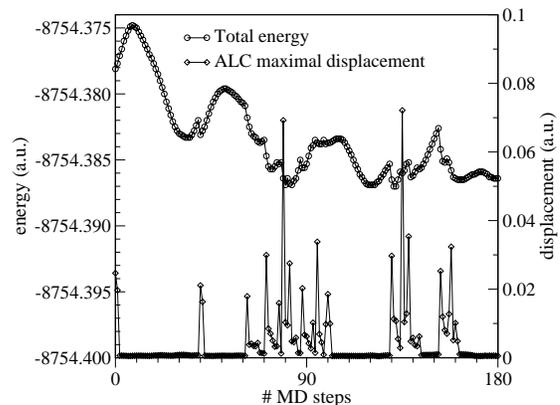

FIG. 2: Total energy and ALC displacements during molecular dynamics simulation of water (512 molecules) at 300 K with localization radius of 9 Bohr
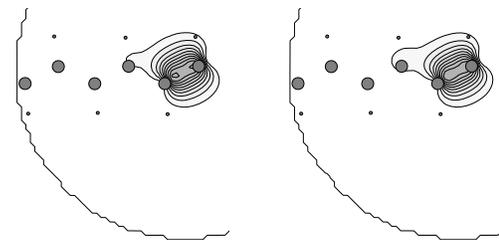

FIG. 3: Contour plot of one localized obitals for two ground state local minimas for polyacetylene in the plane of the molecule (MD snapshot at 300 K, $R_c = 11$ Bohr). The right picture corresponds to a state lower in energy by 0.00089 Ha.

a localized orbital corresponding to two ground states local minima for a different system (polyacetylene). These were obtained by minimizing the energy functional from two different starting points.

For an efficient parallelization, we use a standard domain decomposition approach. The discretization grid is evenly divided among the processors. Each processors uses the data associated to its local *subgrid*. Since the localized orbitals spread only over a limited number of these sub-domains, each processor has to deal with only a fraction of the total number of orbitals. When the number of processors is increased proportionally with the size of the system, the amount of data associated to each processor remains constant. This approach is implemented by packing together several non-overlapping localized orbitals into a single *global* array that represents a grid-based function over the whole discretization grid. This packing problem is equivalent to a graph coloring problem. Each localized orbital corresponds to a node and the overlap between two orbitals is represented by an edge joining the two nodes. The graph coloring problem is known to be NP-complete. We solve it approximately using the recursive largest first (RLF) coloring algorithm [23]. This algorithm requires $O(N^2)$ time for our application, but the total time required for this process, carried out at each MD step, is negligible. The values of the
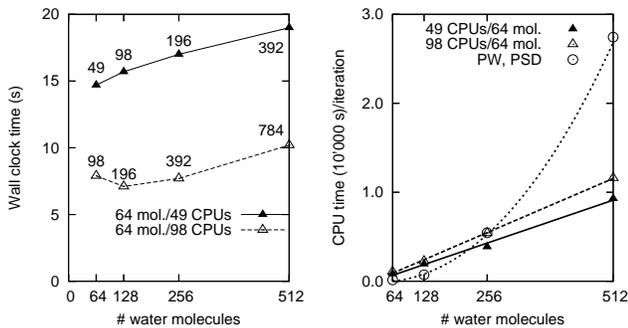
FIG. 4: a) CPU time for 1 SC step with $R_c = 10$ Bohr for liquid water, for 2 ratios of *molecules/CPU* (weak scaling). The numbers next to the data points indicate the number of processors used. b) Comparison with total CPU time required for a preconditioned steepest descent step in PW using a ratio of 35 CPUs/64 molecules.

global arrays at grid points outside of the localization regions are set and kept at zero throughout the calculation.

We evaluated the scaling of our algorithm by measuring timings for one step in the iterative SC solver used to minimize the energy functional. We used liquid water as test system, with up to 512 molecules and 784 CPUs. A strictly O(N) complexity and a perfect scaling would give constant timings for a constant ratio *molecules/CPU*. The results are quite to close to this ideal situation (see Fig. 4). [24] Parallel scaling efficiency is about 80% in this regime. These results are compared with the cost of a standard PW calculations for a state of the art implementation [25] to estimate the crossover between the two methods. Looking at the cost of a PW simple preconditioned steepest descent step, our O(N) approach becomes advantageous above 500-600 atoms. Note that this number does not reflect the efficiency of the iterative algorithms in term of number of iterations required to achieve a good energy conservation. Note also that this is a dense 3D system. For systems with more vacuum, the crossover point would be shifted towards a smaller number of atoms.

In summary, we have demonstrated the possibility of carrying out Born-Oppenheimer MD simulations in O(N) operations using a real-space FD approach with strictly localized orbitals. PW accuracy can be achieved and a comparison with a PW code shows a crossover around 500 atoms for liquid water. Using adaptive localization centers and sufficiently large localization regions are two key elements to ensure accuracy and energy conservation.

[1] S. Bonev, E. Schwegler, T. Ogitsu, and G. Galli, Nature **431**, 669 (2004)).

[2] M. Preuss, W. Schmidt, and F. Bechstedt, Phys. Rev. Lett. **94**, 236102 (2005).

[3] Y.-S. Lee, M. Buongiorno Nardelli, and N. Marzari, Phys. Rev. Lett. **95**, 076804 (2005).

[4] G. Galli, Curr. Opin. Solid State Mater. Sci. **1**, 864 (1996).

[5] F. Mauri and G. Galli, Phys. Rev. B **50**, 4316 (1994).

[6] P. Ordejon, D. A. Drabold, R. M. Martin, and M. P. Grumbach, Phys. Rev. B **51**, 1456 (1995).

[7] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **52**, 1640 (1995).

[8] E. Tsuchida and M. Tsukada, J. Phys. Soc. Jap. **67**, 3844 (1998).

[9] J.-L. Fattebert and F. Gygi, Comput. Phys. Comm. **162**, 24 (2004).

[10] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).

[11] J.-L. Fattebert and J. Bernholc, Phys. Rev. B **62**, 1713 (2000).

[12] J. R. Chelikowsky, N. Trouiller, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994).

[13] E. Hernandez, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **53**, 7147 (1996).

[14] T. Hoshi and T. Fujiwara, J. Phys. Soc. Jpn. **66**, 3710 (1997).

[15] A. Mostofi, C. Skylaris, P. Haynes, and M. Payne, Comp. Phys. Comm. **147**, 788 (2002).

[16] A. Nenciu and G. Nenciu, Phys. Rev. B **47**, 10112 (1993).

[17] E. L. Briggs, D. J. Sullivan, and J. Bernholc, Phys. Rev. B **54**, 14362 (1996).

[18] J. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[19] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[20] M. Fernandez-Serra and E. Artacho, J. Chem. Phys. **121**, 11136 (2004).

[21] W. Silvestrelli and M. Parrinello, Phys. Rev. Lett. **82**, 3308 (1999).

[22] J. Grossman, E. Schwegler, E. Draeger, F. Gygi, and G. Galli, J. Chem. Phys. **120**, 300 (2004).

[23] F. Leighton, J. res. Nat. Bur. Standard **84**, 489 (1979).

[24] Timings measured on the Lawrence Livermore National Laboratory Linux cluster "Thunder", a 1024 nodes computer, with 4 Intel Itanium 1.4 GHz CPUs/node and Qs-Net Elan interconnect.

[25] F. Gygi, QBOX: A first-principles molecular dynamics simulation program, Lawrence Livermore National Laboratory, 2005