

The linearly scaling 3D fragment method for large scale electronic structure calculations

Zhengji Zhao¹, Juan Meza², Byounggak Lee³, Hongzhang Shan², Erich Strohmaier², David Bailey², and Lin-Wang Wang²

¹ National Energy Research Scientific Computing Center (NERSC)

² Computational Research Division, Lawrence Berkeley National Laboratory

³ Physics Department, Texas State University

ZZhao@lbl.gov

Abstract. The Linearly Scaling three-dimensional fragment (LS3DF) method is an $O(N)$ *ab initio* electronic structure method for large-scale nano material simulations. It is a divide-and-conquer approach with a novel patching scheme that effectively cancels out the artificial boundary effects, which exist in all divide-and-conquer schemes. This method has made *ab initio* simulations of thousand-atom nanosystems feasible in a couple of hours, while retaining essentially the same accuracy as the direct calculation methods. The LS3DF method won the 2008 ACM Gordon Bell Prize for algorithm innovation. Our code has reached 442 Tflop/s running on 147,456 processors on the Cray XT5 (Jaguar) at OLCF, and has been run on 163,840 processors on the Blue Gene/P (Intrepid) at ALCF, and has been applied to a system containing 36,000 atoms. In this paper, we will present the recent parallel performance results of this code, and will apply the method to asymmetric CdSe/CdS core/shell nanorods, which have potential applications in electronic devices and solar cells.

1. Introduction

Nanostructures have wide applications in biological imaging, light emitting diodes, solar cells, and other electronic devices. The sizes of the nanostructures are so small that they have very different electronic and optical properties from their bulk counter parts, due to quantum confinement effects. Nevertheless, to study the properties of nanostructures, one needs to do *ab initio* calculations on systems containing 1,000 to 100,000 atoms, which are too large for direct *ab initio* methods. Direct methods have been applied to systems with 2,000 atoms at most [1]. This is because even the simplest *ab initio* methods, e.g., the density functional theory (DFT) methods under the local density approximation (LDA), are computationally expensive, scaling as $O(N^3)$, where N is the size of the system. In addition, due to communication bottlenecks the parallelization of direct LDA methods is limited to roughly 10,000 processors [1]. In fact, the most widely used direct LDA code, VASP, is difficult to scale to even several thousand processors. Therefore, both the computational costs and the limit on parallelization call for a new approach.

Over the past decade, many $O(N)$ methods have been developed [2]. These approaches can be classified into three main categories: the local orbital methods [3], the truncated D-matrix methods [4],

and divide-and-conquer methods [5]. While these methods have successfully reduced the computational cost and have been applied to large systems, several fundamental technical issues remain that are difficult to overcome. For example, in the commonly used local orbital methods, there exist extraneous local minima in the total energy functional, which make the total energy minimization difficult (convergence problem). This is due to constraining the wavefunctions on the local orbital manifold. Moreover, the overlap between neighboring local orbitals makes it difficult to scale these methods to large numbers of processors. In short, for the previous $O(N)$ methods, especially the local orbital methods and the truncated D-matrix methods, on top of some fundamental technical issues, a main challenge is to scale the codes to tens of thousands of computer processors while preserving the *ab initio* accuracy.

Recently we have developed a new $O(N)$ method, the linearly scaling three-dimensional fragment (LS3DF) method [6,7]. It is a divide-and-conquer approach that scales to tens of thousands of computer processors quite easily, and yields essentially the same results as direct LDA methods. The LS3DF method won the 2008 ACM Gordon Bell Prize for algorithm innovation [8]. In this paper, we will briefly describe the LS3DF method, focusing on its recent parallel performance results, then describe how we have used LS3DF to study the electronic structures of asymmetric CdSe/CdS core/shell nanorods, which have potential applications in solar energy conversions.

2. The LS3DF method

The LS3DF method is based on the short-ranged nature of quantum mechanical effects. The total energy of a system can be split into the classical electrostatic energy and the quantum mechanical energy (kinetic energy and exchange correlation energy). The electrostatic interaction is long-ranged, therefore the electrostatic energy must be calculated by solving a Poisson equation globally. But the quantum mechanical effects are short-ranged, therefore they can be solved locally, and the total quantum mechanical energy for the system can be obtained by combining the locally calculated quantum energies. In our LS3DF method, we divide a large system into small pieces (fragments), and independently calculate a solution on each fragment. We then patch the local solutions together to obtain the total energy and the total charge density for the whole system. The heart of the LS3DF method is a novel patching scheme that effectively cancels out the artificial boundary effects. Figure 1 and 2 illustrate our division and patching scheme using a 2D example for simplicity. In Figure 1, a periodic super cell is divided into 4×4 pieces. At each grid point (i,j) , we introduce 4 fragments (along the right-upper direction) with different sizes, 1×1 , 2×1 , 1×2 and 2×2 . All fragments at all grid points (i,j) ($i=1, \dots, 4; j=1, \dots, 4$) will be calculated independently using a direct LDA method, e.g., PEtot [9], a planewave pseudo potential LDA code. The fragments will then be summed up according to the patching scheme illustrated in Figure 2. Where the 1×1 (red) and 2×2 (blue) fragments are positive fragments, and the other two, the 1×2 (yellow) and the 2×1 (green) are negative fragments.

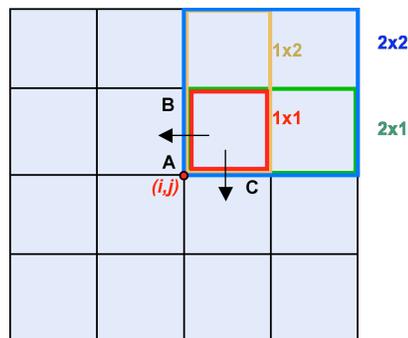


Fig. 1. A schematic view of the division of a system into small fragments. This 2D periodic super cell is divided into 4×4 fragment grids. At each fragment grid point (i,j) , 4 fragments with different sizes are introduced. Where the red, green, yellow, and blue rectangles represent the fragments of size 1×1 , 2×1 , 1×2 and 2×2 , respectively.

$$\text{Total} = \sum_{(i,j)} \left\{ \boxed{} - \boxed{} - \boxed{} + \boxed{} \right\}$$

Fig. 2. The schematic view of the fragment patching scheme in the LS3DF method for 2D systems. Here the yellow (1×2) and the green (2×1) fragments are negative fragments, and blue (2×2) and red (1×1) are positive fragments.

We can illustrate how the patching scheme yields the original system from the fragments. Let's consider the area covered by the red square in Figure 1. For convenience, we denote the fragment 1×1 introduced at the grid point (i,j) as $F_{11}(i,j)$. By counting how many positive and negative fragments cover this area, one can easily see whether this area is described properly after all the fragments are added up. This area is covered by 5 positive fragments, they are $F_{11}(i,j)$, $F_{22}(i-1,j-1)$, $F_{22}(i,j-1)$, $F_{22}(i,j)$ and $F_{22}(i-1,j)$. And this area is also covered by four negative fragments, which are $F_{21}(i,j)$, $F_{21}(i-1,j)$, $F_{12}(i,j)$ and $F_{12}(i,j-1)$. When these fragments are summed up using the patching scheme in Figure 2, the red square area is covered only once after 4 positive and 4 negative fragments cancel out. We can also show the artificial boundary is removed in this patching scheme. Let's consider the left boundary of the red square (edge AB). We can define a direction (outward) for this boundary as shown with a left arrow in Figure 1. We can count how many fragments contain this boundary. They are three positive fragments, $F_{11}(i,j)$, $F_{22}(i,j)$, and $F_{22}(i,j-1)$ and three negative fragments, $F_{12}(i,j)$, $F_{12}(i,j-1)$, and $F_{21}(i,j)$. When these six fragments are summed up, the edges from the three negative fragments cancel out the edges from the other three positive fragments. Similarly we can see the artificial corners (e.g., the corner BAC, outward direction) cancel.

The patching scheme for 2D systems can be generalized to 3D systems straightforwardly. In 3D cases, at each grid point (i,j,k) , eight fragments (along the right-upper direction) with different sizes are introduced, they are four positive fragments, $2 \times 2 \times 2$, $2 \times 1 \times 1$, $1 \times 2 \times 1$ and $1 \times 1 \times 2$, and four negative fragments, $2 \times 2 \times 1$, $2 \times 1 \times 2$, $1 \times 2 \times 2$, and $1 \times 1 \times 1$. The patching scheme is

$$\text{Total} = \sum_{i,j,k} \{F_{222} + F_{211} + F_{121} + F_{112} - F_{221} - F_{212} - F_{122} - F_{111}\}$$

The boundary effect cancellation assumes that the fragments that share a given boundary have very similar charge densities near that boundary. Our tests have shown that this assumption always holds as long as the smallest fragment ($1 \times 1 \times 1$) is not too small and the fragments have a band gap (after some artificial boundary passivations). When a typical 8-atom unit cell is chosen as the $1 \times 1 \times 1$ fragments, the errors of the total energy, charge density, atomic force and the dipole moment are well below the typical errors introduced by other approximations of the numerical calculations (e.g., pseudopotentials, planewave basis cut offs). Thus the LS3DF method gives essentially the same results as the direct methods. In contrast to the convergence issues that is common in other $O(N)$ methods, the self consistent (SC) iterations in LS3DF converges at a comparable rate as the direct methods. For a system containing a few thousands atoms, the total energy typically converges to 10^{-6} a.u. within 50-60 SC charge density mixing iterations. The LS3DF method outperforms the direct LDA methods when the system contains more than 550 atoms. For nanostructures with 10,000 atoms, the LS3DF would be

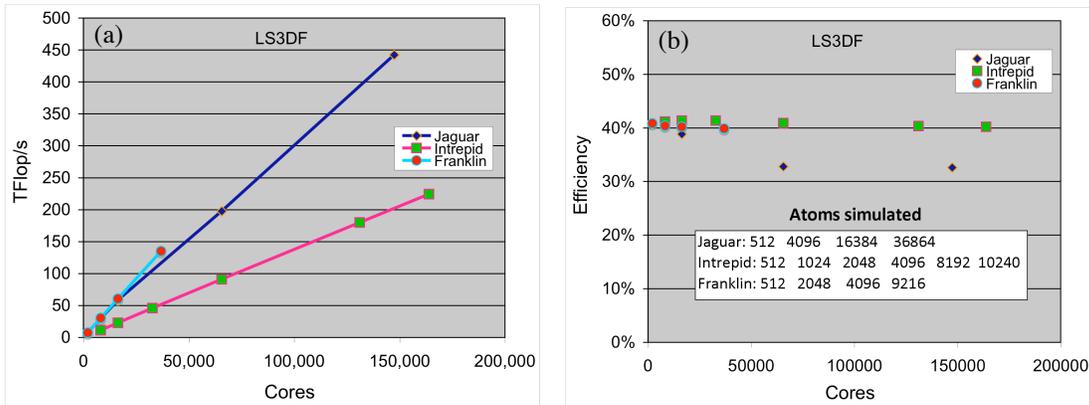


Fig.3. Weak scaling floating point operation rates (a) and the computational efficiency (b) of the LS3DF method on different machines. The systems used were $\text{ZnTe}_{1-x}\text{O}_x$ alloy ($x=8\%$) with various number of atoms. The planewave cut off energy was 60 Ry. The flops were measured for one SCF iteration step. The floating point operations were measured in the double precision, using the profiling tool Craypat 4.1.

faster by 3 orders of magnitudes, even presuming the direct LDA methods can scale up to thousands of processors. For more details, see [6,7].

One of the great advantages of the LS3DF method is its excellent parallel scaling. We have run our code on the Cray XT4 (Franklin) at NERSC, the Cray XT5 (Jaguar) at OLCF, and the Blue Gene/P (Intrepid) at ALCF. Figure 3 shows the weak scaling results on these three machines, indicating that LS3DF exhibits nearly linear scaling up to the maximum available processor cores on all three systems. It achieved 135 Tflop/s on 36,864 cores on Franklin at 40% efficiency; 224 Tflop/s on 163,840 cores on Intrepid at 40% efficiency; 442 Tflop/s on 147,456 cores on Jaguar at 33% efficiency. Additional LS3DF performance results are given in [8].

3. Electronic structure calculations for asymmetric CdSe/CdS core/shell nanorods

With the advance of synthetic methods, numerous nanostructures with different shapes have been synthesized in research laboratories. Recently Carbone and his colleagues have synthesized asymmetric core/shell structures, using the newly developed seed growth method [10]. In the asymmetric core/shell nanorods, a CdSe core is embedded at one end of a CdS quantum rod shells. By changing the sizes of the core and the shell (both diameter and length), one can manipulate the electronic structures inside the nanorods. Hence these nanostructures appear to be particularly interesting for solar cell applications. In addition, these asymmetric core/shell structures provide a system on which one can study the quantum confinement effect, the band alignment, the strain (due to the core/shell lattice constant mismatch), and the surface effects.

We have applied the LS3DF method to the asymmetric CdSe/CdS core/shell nanorods with both the Cd atoms terminated and with the Cd+S atoms terminated surfaces, as well as to their counterparts:

the pure CdS nanorods, to study how the core and the surface affect the electronic structures inside these nanorods. The LS3DF calculations give us the total charge density and the total potential. Using the resulting potential, we have calculated the valence band maximum (VBM, hole) and the conduction band minimum (CBM, electron) states and the band gaps (Table 1), utilizing the folded spectrum method [11]. We have also calculated the dipole moments (Table 2) of these nanorods. One can see that in the nanorods with the Cd terminated surface, the introduction of the CdSe core results in a significant change in the band gaps (0.24 eV) and dipole moment (-9.65 a.u.), while in the nanorods with the Cd+S terminated surface, the role of the CdSe core seems to be surprisingly small, both on the band gap and the dipole moment. Figure 4 shows the charge density isosurface of the hole and the electron states of the four nanorods. One can see that in the nanorods with the Cd terminated surface, the introduction of the CdSe core significantly changes the localization of the hole states, while in the nanorods with the Cd+S terminated surface, the hole state localization seems not be affected by the presence of the CdSe

	Cd termin. (a.u.)	Cd+S termin. (a.u.)	Dipole mom. change (a.u.)
Pure CdS nanorods	-15.623	-28.415	-12.792
CdSe/CdS core/shell	-25.277	-28.108	-2.830
Dipole mom. change (a.u.)	-9.654	0.307	

Table 2. The calculated dipole moments along the c-axis of the four CdS nanorods (the components of the dipole moments in the other two directions are small, not shown here). The dipole moment changes due to the different surfaces (column 4), and the CdSe cores (row 4) are also shown in the table.

	Cd termin. (eV)	Cd+S termin. (eV)	Band gap change (eV)
Pure CdS nanorods	2.655	2.498	-0.156
CdSe/CdS Core/shell	2.415	2.403	-0.011
Band gap change (eV)	-0.240	-0.095	

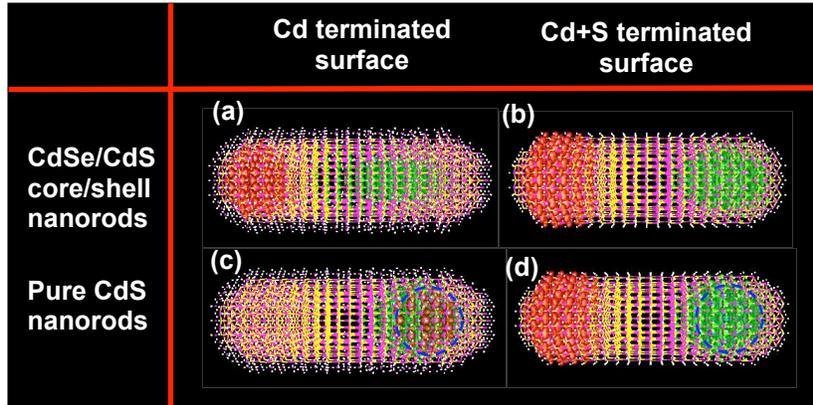
Table 1. The calculated band gaps of the four nanorods. The band gap changes due to the different surfaces (column 4) and the CdSe core (row 4) are also shown in the table.

core, indicating the Cd+S terminated surface has the dominant effects to the hole localizations. Further study on the surface passivation effect is under way [12].

4. Conclusions

We have briefly described the LS3DF method for *ab initio* electronic structure calculations. This scheme features $O(N)$ scaling in terms of system size, and also scales very well in parallel performance. We presented parallel scaling results, summarized the accuracy and the SCF convergence rate, and then applied this method to analyze asymmetric CdSe/CdS core/shell nanorods. We demonstrated that the scheme scales linearly to hundreds of thousands of computer processors; it yields essentially the same results as the direct LDA methods; it can calculate a nanosystem with thousands of atoms self-consistently in a couple of hours. We expect wide applications of LS3DF method in nanostructure calculations.

Figure 17. Isosurface of the charge densities of the conduction band minimum (CBM, green) and the valance band maximum (VBM, red) states of the four CdS nanorods with/without CdSe core. Where (a) and (b) are for the pure CdS nanorods with the Cd terminated and the Cd+S terminated surfaces, while (c) and (d) are for the corresponding CdSe/CdS core/shell nanorods, respectively. The isovalue larger than 0.001 e/bhor^3 was shown for both VBM and CBM states. The blue dashed circle shows the CdSe core area. These nanorods are constructed as a wurzite structure, and have 2.8 nm in diameter, and 8.4 nm in length (c-axis). The diameter of the CdSe core is 2.1 nm. There are 3063 and 2298 atoms in the nanorods with the Cd terminated and the Cd+S terminated surfaces, respectively. Where the magenta, yellow, and blue dots represent the Cd, S, and Se atoms, respectively, and the white dots represent pseudo H atoms which passivate the surface dangling bonds.



5. Acknowledgments

This work was supported by the ASCR and the BES offices in the DOE, Office of Science, under LAB03-17 initiative, contract No. DE-AC02-05CH11231. It was supported by the DOE INCITE project. It used the resources of National Energy Research Scientific Computing Center (NERSC), Oak Ridge Leadership Computing Facility, and Argonne Leadership Computing Facility (ALCF). The authors would like to thank Jeff Larkin at Cray Inc., and Katherine Riley at ALCF for their help with running the LS3DF code on Jaguar and Intrepid, respectively. They would also thank Ying Luo at Beijing Normal University for providing the VFF relaxed core/shell nanostructures.

References

- [1] F. Gygi, *et al*, Proceedings of the 2005 ACM/IEEE Conference on Supercomputing, (2005).
- [2] G. Goedecker, Rev. Mod. Phys. 71:1085, 1999.
- [3] G. Galli, *et al*, Phys. Rev. Lett. **69**, 3547(1992); J.-L.Fattebert, *et al*, Phys.Rev.B **73**,115124 (2006)
- [4] X.P. Li, *et al*, Phys. Rev. B **47**, 10891 (1993); D.R. Bowler, *et al*, phys. stat. sol. b **243**, 989 (2006).
- [5] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
- [6] L. W. Wang, Z. Zhao, and J. Meza, Phys. Rev. B **77**, 165113 (2008).
- [7] Z. Zhao, J. Meza, and L. W. Wang, J. Phys.: Condens. Matter **20** (2008) 294203.
- [8] L. W. Wang, *et al*, ACM Gordon Bell submission, Supercomputing (2008).
- [9] <http://hpcrd.lbl.gov/linwang/PEtot/PEtot.html>
- [10] L. Carbone, *et al*, Nano Lett., **7**(10) 2942-2950 (2007)
- [11] L. W. Wang and A. Zunger, J. Chem. Phys. **100**, 2394 (1994).
- [12] Z. Zhao, and L. W. Wang, *Application of the LS3DF method in CdSe/CdS core/shell nanorods*, to be submitted.