Linear-scaling three-dimensional fragment method for large-scale electronic structure calculations

Lin-Wang Wang, Zhengji Zhao,[*] and Juan Meza

Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

We present a linear scaling *ab initio* total energy electronic structure calculation method, which is simple to implement, easily to parallelize, and produces essentially the same results as the direct *ab initio* method, while it could be thousands of times faster. Using this method, we have studied the dipole moments of CdSe quantum dots, and found both significant bulk and surface contributions.

PACS number(s): 71.15.Ap, 73.22.−f

I. INTRODUCTION

In the past decade, we have witnessed an increasing number of experimental investigations of structural, electronic, and optical properties of ever more complex nanostructures. This trend calls for a corresponding set of theoretical *ab initio* calculations on these nanosystems, some of which may contain tens of thousands of atoms. However, due to the $O(N^3)$ computational scaling[1] of the direct density functional theory (DFT), it can only be applied to about one to two thousand atoms despite the ever increasing computer powers and parallelism.[2] Over the past 15 years, many linear scaling $O(N)$ electronic structure algorithms have been proposed.[3] A common algorithm is based on localized orbitals. Unfortunately, the use of localized orbitals can lead to local minimum in the energy functionals, causing convergence problem in the calculations.[4] Besides, it is difficult to represent the local orbitals with plane wave basis which is widely used in materials science simulation. The overlaps between neighboring local orbitals also make the code parallelization not so straightforward. Overall, there is a continuous need for new and simple $O(N)$ *ab initio* methods which can be used by a wider population in the computational electronic structure community.

In this paper, we present a new $O(N)$ *ab initio* electronic structure method and use it to study dipole moments in CdSe quantum dots. This method satisfies the following criteria for a good modern $O(N)$ algorithm: (1) it is accurate, obtaining essentially the same results compared to the direct *ab initio* method; (2) it is simple, which makes it easy to be implemented from an existing *ab initio* code; (3) it is trivially parallelizable, which makes it suitable for large-scale computation; (4) it is applicable to any *ab initio* method, not restricted to DFT.

II. FORMALISM

Our method is based on the observation that the total energy of a given system can be split into two parts: the electrostatic energy and the quantum mechanical energy (e.g., the kinetic energy and exchange correlation energy). While the electrostatic energy is long range and must be solved via a global Poisson equation, the computationally expensive quantum mechanical energy is short range[5] and can be solved locally. Our idea is to divide the whole system into small fragments, calculate the quantum mechanical energies of these fragments, and then sum the separate fragment energies to obtain the energy of the whole system. The core of our algorithm is a patching scheme that sums the fragment energies in such a way that the artificial boundary effects caused by the division of the system essentially cancel out.

Our division and patching scheme is illustrated in Fig. 1, which uses a two-dimensional system for clarity. In Fig. 1, a periodic supercell is divided into $m_1 \times m_2$ small pieces. From each grid corner (i,j), we can define four fragments, with their sizes (S) equal to (in units of the smallest piece) $S=1 \times 1$, 1×2, 2×1, and 2×2, respectively. Suppose we calculate the quantum energy $E_{i,j,S}$ and charge density $\rho_{i,j,S}$ of all of these fragments. Then the total quantum energy of the system can be computed by $E=\Sigma_{i,j,S}\alpha_S E_{i,j,S}$ and the total charge density by $\rho(r)=\Sigma_{i,j,S}\alpha_S\rho_{i,j,S}(r)$. Here, $\alpha_S=1$ for the $S=1 \times 1$ and 2×2 fragments, and $\alpha_S=-1$ for the $S=1 \times 2$ and 2×1 fragments.[6] By using the above summation, the long surface [the edge of $(i,j)-(i+2,j)$ in Fig. 1] of the 2×1 fragment will cancel the same surface of the 2×2 fragments, and the surface of the 1×1 fragment will cancel out the new unwanted short surfaces [the edge of $(i,j)-(i,j+1)$] of the 2×1 fragment. Such cancellation is complete for all the surfaces (edges in Fig. 1) and corners after the summation $\Sigma_{i,j,S}\alpha_S$ is carried out for all i, j, and S in the total energy and charge expression.
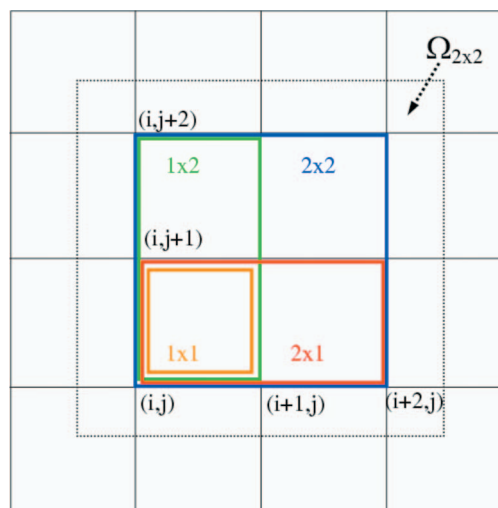

FIG. 1. (Color) A schematic view of the division of the space into fragments. The dashed line denotes the region of the 2×2 fragment.

The above scheme can be extended to a three-dimensional system in a straightforward manner. Here, a periodic supercell is divided into $m_1 \times m_2 \times m_3$ fragments, and from each grid point corner (i,j,k) we can define eight fragments, with sizes $S(\alpha_S)$ equal to $1 \times 1 \times 1 (-)$, $1 \times 1 \times 2 (+)$, $1 \times 2 \times 1 (+)$, $2 \times 1 \times 1 (+)$, $1 \times 2 \times 2 (-)$, $2 \times 1 \times 2 (-)$, $2 \times 2 \times 1 (-)$, and $2 \times 2 \times 2 (+)$. Now, $E = \Sigma_{i,j,k,S}\alpha_S E_{i,j,k,S}$. This formula has the same property of canceling out all the surface, edge, and corner effects.

We can now compare our linear scaling three-dimensional fragment (LS3DF) method to other divide-and-conquer approaches. In the method proposed by Yang,[7,8] the spatial domain is divided into overlapping subdomains. A positive spatial partition function $P(r)$, where $P(r)$ equals 1 near the center of the subdomain and gradually goes to 0 near the boundary, is used for patching the subdomains. The idea is to use only the central part of each subdomain, instead of using boundary effect cancellations as in our method. The use of the partition function $P(r)$, however, creates some technical problems including how to partition the kinetic energy,[8] how to express the total energy in a variational form and how to maintain charge neutrality. Note that by reducing the size of our $2 \times 2 \times 2$ fragments,[6] we can also reduce the overlap between fragments, much as in the method proposed by Yang.[7,8]

We can also compare our LS3DF method with the fragment molecular orbital (FMO) method.[9,10] FMO is specifically designed for biological systems where a long chain molecule is divided into many small pieces (monomers). In the FMO method, all of the monomer and monomer-monomer pairs are calculated to take into account the covalent bonds that are broken in the subdivision. In contrast, we have three-dimensional fragments with different sizes in a spatially compact form. If we identify our smallest $1 \times 1 \times 1$ fragment with the monomers in FMO, then we calculate up to eight monomer clusters (the $2 \times 2 \times 2$ fragments). In summary, LS3DF uses regular spatial location and division and has a rigorous boundary effect cancellation, unlike FMO. In fact, as we will show below, the error in LS3DF will drop rapidly as the fragment size increases.

In our implementation of the LS3DF method, we start with a 3D periodic supercell that is divided into an $M = m_1 \times m_2 \times m_3$ grid. Each atom is assigned to one fragment depending on its spatial location (which square it falls into in Fig. 1). The artificially created surfaces of the fragments are passivated with hydrogen or partially charged pseudohydrogen atoms to fill the dangling bonds.[11] We will denote the fragment wave function as $\psi_{F,i}(r)$, where i is the wave function index, and $F = (i,j,k,S)$ is the index for the fragment. Note that $\psi_{F,i}(r)$ is only defined within its own fragment's spatial domain Ω_F, which is of size S plus a surface buffer region as indicated by the dashed line in Fig. 1. We can now write the total energy E_{tot} of the system as a variational expression in terms of the fragment wave functions $\psi_{F,i}(r)$ (for zero temperature calculation):

$$E_{tot} = 2\sum_F \alpha_F \sum_{i=1,N_F/2} \int \psi_{F,i}^*(r)\left[-\frac{1}{2}\nabla^2\right]\psi_{F,i}(r)d^3r$$
$$+ V_{ion}(r)\rho_{tot}(r)dr + \frac{1}{2}\int \frac{\rho_{tot}(r)\rho_{tot}(r')}{|r-r'|}d^3rd^3r'$$
$$+ \int \epsilon_{xc}[\rho_{tot}(r)]\rho_{tot}(r)d^3r + \sum_F \alpha_F \int \Delta V_F(r)\rho_F(r)d^3r,$$
$$(1)$$

where $\rho_{tot}(r) = \Sigma_F\alpha_F\rho_F(r)$, $\alpha_F = \alpha_S$, and the fragment charge density $\rho_F(r) = 2\Sigma_{i=1,N_F/2}|\psi_{F,i}(r)|^2$, where N_F is the total number of electrons in fragment F after passivation. $V_{ion}(r)$ in Eq. (1) is the total ionic potential. The term $\Delta V_F(r)$ is an additional surface passivation that is only nonzero near the boundary of the fragment. For different fragments sharing a same boundary B, their $\Delta V_F(r)$ at that boundary B should be the same. Due to the fragment cancellations, the net value of the last term in Eq. (1) should be small. The amplitude of this term can be used as a measure of the accuracy for this method.

The total energy E_{tot} is a variational minimum (or maximum, depending on the sign of α_F) with regard to $\psi_{F,i}(r)$, subject to the orthonormal constraints: $\int_{\Omega_F}\psi_{F,i}^*(r)\psi_{F,j}(r)dr = \delta_{i,j}$. Thus, we can derive the fragment Kohn-Sham equation from $\delta E_{tot}/\delta\psi_{F,i}^*(r) = 2\alpha_F\epsilon_{F,i}\psi_{F,i}(r)$, which results in

$$\left[-\frac{1}{2}\nabla^2 + V_F(r)\right]\psi_{F,i}(r) = \epsilon_{F,i}\psi_{F,i}(r), \quad (2)$$

where

$$V_F(r) = V_{tot}(r) + \Delta V_F(r) \quad \text{for } r \in \Omega_F, \quad (3)$$

and $V_{tot}(r)$ is the usual local density approximation (LDA) total potential calculated from ρ_{tot} by solving a global Poisson equation for the whole system. Of practical importance is the observation that the calculations in Eq. (2) can be carried out independently for each fragment making the computation trivially parallel. The charge density self-consistency can be achieved iteratively using the usual potential mixing scheme[1] for V_{tot}. Due to the variational principle, atomic forces can be calculated using the Hellman-Feyman theory. To compute the surface passivation potential, we have used the atomic charge densities to construct a $\rho_{F,atom}(r)$, $\rho_{tot,atom}(r)$, and $V_{tot,atom}(r)$. From $\rho_{F,atom}(r)$, we have also calculated a $V_{F,atom}(r)$ using a LDA formula for the fragment. We then have $\Delta V_F(r) = V_{F,atom}(r) - V_{tot,atom}(r)$. To assure that $\Delta V_F(r)$ at a given boundary B is the same for fragments sharing this common boundary, we take the average among all of the fragments sharing this boundary. With the vacuum buffer region, we have used a plane wave expansion for the wave functions $\psi_{F,i}(r)$ and norm conserving pseudopotentials for the Hamiltonian. Equation (2) is solved using a conjugated gradient method based on the plane wave code, PETOT.[12]

III. NUMERICAL TEST

We first compare the LS3DF method against the direct LDA method for a $Si_{235}H_{104}$ quantum dot (QD) with surface

TABLE I. The convergence of the LS3DF results compared with direct LDA results for bulk Si calculations. The fragment sizes $0.5a$, $1a$, and $1.5a$ correspond to 8, 64, and 216 Si atoms in the $2 \times 2 \times 2$ fragments, respectively. ΔE is the total energy error, and $\Delta \rho$ is the total charge density error.

Fragment size	$0.5a$	$1a$	$1.5a$
ΔE (meV/at)	30	2.9	4.0
$\sum_F \alpha_F \int \Delta V_F \rho_F dr$ (meV/at)	213	5.5	1.0
$\Delta \rho$	1.1%	0.14%	0.08%

hydrogen passivation and a plane wave basis set cutoff of 35 Ry. The smallest $1 \times 1 \times 1$ fragment is an eight atom unit cell. With these parameters, the total energy difference between the LS3DF method and the direct LDA method is 3 meV/atom. The average charge density difference is 0.2%, while the average atomic force difference is 5×10^{-5} a.u. We also tested Si slabs and rods, and CdSe quantum dots (where an internal electric field exists). The errors for those tests are similar to the above Si QD results. To test the fragment size effects more systematically, we have calculated the bulk Si with LS3DF method. Our results, shown in Table I, indicate that the errors drop rapidly as the fragment size increases from $0.5a$ to $1.5a$ (a is the lattice constant of bulk Si). We note that the total energy error does increase a bit in going from $1.0a$ to $1.5a$, however. This is due to the use of negative fragments, which allows the computed energy to approach the exact result from both above and below. A more robust number is the charge density error, which drops rapidly with the fragment size. We also calculated the quantum dot polarization under an external electric field in a Si quantum dot using the $1.0a$ fragment size. The LS3DF and direct LDA differences for the response charge and total induced dipole moment are both about 2%. Thus, we claim that, with a $1.0a$ fragment size, the LS3DF method is accurate enough for most practical calculations.

Figure 2 shows the convergence of the self-consistent iterations. One can see that the LS3DF method has a convergence rate similar to the direct LDA method, avoiding the convergence problems seen in some of the other $O(N)$ methods.[4] Since each fragment Kohn-Sham equation is solved independently, the parallelization of this method is
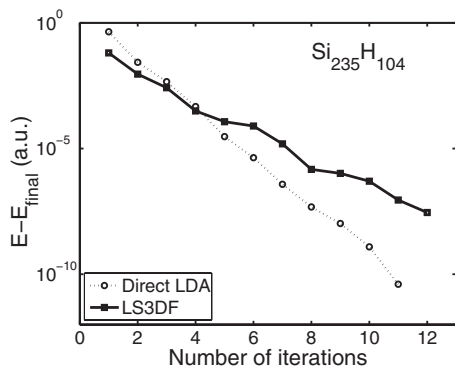

FIG. 2. Self-consistent convergence curves for LS3DF and direct LDA method.

straightforward. The time spent on the global Poisson equation is only a fraction of the total computational time. As a result, we have been able to achieve an excellent (up to 80%) linear scaling with the number of processors (up to 8000 processors in our tests). Based on actual computer total floating point operations counts, the crossover size of the linear scaling LS3DF method and the $O(N^3)$ scaling direct LDA method is around 500 atoms with the $1.0a$ fragment size. To demonstrate the power of the LS3DF method, we have calculated a 15 000 atom Si quantum dot, which took 30 min for one self-consistent iteration using 2048 processors on an IBM SP Power3 machine. By comparison, if a direct LDA method had been used, it would have taken weeks using similar number of processors.[13]

The reported crossovers with direct LDA calculation for localized orbital and density matrix methods are about 500 atoms. Since these crossovers are similar to ours for similar accuracy, we can deduce that the LS3DF method should be as efficient as those $O(N)$ methods.[4] Alternatively, one can also estimate the computational cost (for a system with $2N$ electrons) as follows. First, note that for an accurate LS3DF calculation, it is not necessary to have small quantum confinement effects for the fragments. It is only necessary for one side of the fragment to have a small effect on the opposite side in terms of density and kinetic energy density. Such effects should have a similar decay length as the localized Wannier function, because the densities can be expressed as the sum of the Wannier function squared. In practice, we do find that the 64 atom $2 \times 2 \times 2$ fragment size is similar to the orbital size in the localized orbital method for similar accuracies.[4] Most of the LS3DF computational cost is in the computation of the $2 \times 2 \times 2$ fragments. There are $M = m_1 \times m_2 \times m_3$ such fragments, each with $16N/M$ electrons. Thus, in total, there will be $8N$ fragment wave functions each in a domain of $\Omega_{2 \times 2 \times 2}$. In the localized orbital method, the number of localized wave functions is about N to $2N$ depending on the implementation.[4] Our method can be as efficient as the localized orbital method because the iterative convergence of the fragment wave functions is fast due to the wave function decoupling of the fragments, and the $O(N^3)$ step for each fragment calculation is not dominating for the fragment sizes we are using.

We next study the total dipole moments of CdSe quantum dots, which only becomes computationally amenable due to the development of our LS3DF method. This is a longstanding physics problem that stems back more than ten years.[14] Experimentally, it was found that not only there are dipole moments for wurtzite structure CdSe quantum dots and rods,[15,16] there are also similar magnitude dipole moments for zinc blende structure ZnSe quantum dots[15] (which should not have a bulk contribution due to the zinc blende symmetry). As such, there has been a controversy regarding the cause of the measured dipole moment: intrinsic bulk dipole[16] or surface passivation?[15]

We first calculated the dipole moments for a small 178 atom wurtzite CdSe quantum dot. Using a $1 \times 1 \times 1$ fragment of 12 Cd+Se atoms, the LS3DF z-direction (c-axis) dipole
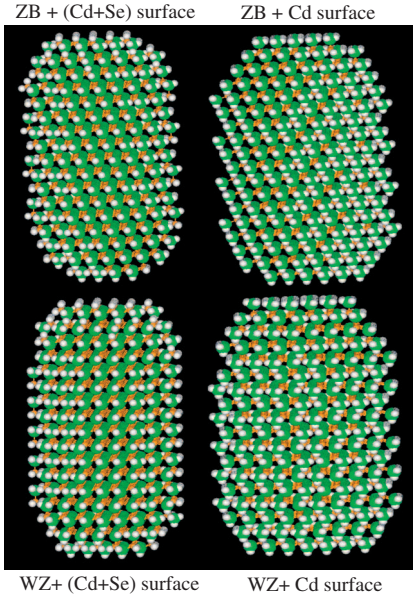
FIG. 3. (Color) The four calculated CdSe nanorods. The green atoms are Cd, yellow atoms are Se, and white atoms are surface hydrogen.

moment was computed to be 3.52 a.u., while the LDA result was 3.49 a.u. The absolute difference (0.03 a.u.) is much smaller than the error introduced by using different pseudopotentials. In the dipole moment calculation, the Poisson equation is solved using an open boundary condition, instead of a periodic boundary condition for the whole system, so there are no neighboring dipole-dipole interactions.

We have calculated four similarly sized elongated CdSe quantum dots (rods) containing a few thousand atoms, as shown in Fig. 3. Two dots are in a wurtzite (WZ) structure and two are in a zinc blende (ZB) structure. Experimental crystal structures and internal atomic coordinates are used for the CdSe WZ structure. Due to the lack of experimental surface passivation details at the atomic level, we used pseudohydrogen atoms to fill the surface dangling bonds, representing an ideal passivation situation.[11] These surface H atoms are placed at the center positions of the cutoff bonds. To study the effects of different surface passivations, we chose two types of surface models, one with both Cd and Se atoms on the surface (Cd+Se terminated) and the other with only Cd atoms on the surface (Cd terminated).

TABLE II. The total z- (c-axis) direction dipole moments of CdSe quantum rods (the x,y dipole components are very small). The total atomic number N_{atom} includes the numbers of Cd, Se, and surface H atoms.

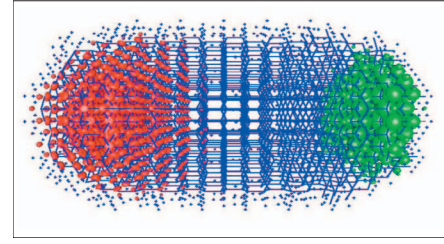| QD	N_{atom}	Struct.	Termin.	Dipole (a.u)
$Cd_{954}Se_{718}$	2616	ZB	Cd	−13.1
$Cd_{961}Se_{724}$	2633	WZ	Cd	71.0
$Cd_{715}Se_{718}$	1955	ZB	Cd+Se	−21.5
$Cd_{714}Se_{724}$	1956	WZ	Cd+Se	51.8

FIG. 4. (Color) The electron (red) and hole (green) states shown in their isosurfaces with an isovalue of $0.0002e/bohr^3$ in the $Cd_{961}Se_{724}$ wurtzite structure quantum rods with Cd atom terminations.

Our final results are listed in Table II. We can see that the total dipole moment depends sensitively on the crystal structure. For ZB structure, the dipole moment is negative, while it is positive for WZ structure. Due to the symmetry, ZB does not have bulk dipole contribution. So the dipole moments of the ZB quantum rod structures must come from the their surface passivations. This effect of the surface contribution is tested by changing the Cd and Se terminated surface to Cd only terminated surface (Fig. 3). We see that this change adds a positive dipole moment to both the ZB and WZ nanorods. Overall, the WZ nanorods have much larger dipole moments. One can attempt to separate the bulk and surface contributions of the WZ dipole moments by assuming that the the surface dipole moments for ZB and WZ nanorods are the same for the same type of surface termination (either Cd termination or Cd+Se termination). Thus, by subtracting the ZB dipole moments from the corresponding WZ dipole moments, we get the bulk dipole contributions of 73.3 and 84.1 a.u. for the $Cd_{714}Se_{724}$ and $Cd_{916}Se_{724}$ WZ quantum dots, respectively. They are exactly proportional to their total number of Cd and Se atoms. The spontaneous polarization bulk dipole moment of WZ structure can been calculated from a bulk WZ/ZB supercell,[17] and we have $P_0 = 0.0143(N_{Cd} + N_{Se})$ a.u. Using this formula, the total bulk contributions for the above two WZ quantum dots should be 20.5 and 23.5 a.u., respectively. Curiously, these estimated bulk contributions are about 3.6 times smaller than the direct calculated results. Further work is under way to explain this difference.

Finally, to study the effect of the dipole moment on the internal electronic structure of nanorods, we have taken the potential V_{tot} of Eq. (3) and calculated the band edge electron and hole states of the whole quantum dot using the folded spectrum method.[18] As shown in Fig. 4, the electrons and holes are localized on opposite ends of the rod, indicating the importance of the possible internal electric fields induced by the dipole moments in such quantum dots.[19] We are carrying out further study to investigate this effect.

IV. CONCLUSION

In conclusion, we have presented a divide-and-conquer linear scaling method for *ab initio* total energy calculations. The current method has the following features: (1) it has a

variational formalism, which allows the calculation of atomic forces using Hellman-Feyman theory; (2) it is very accurate, with a total energy error of about a few meV/atom; (3) it is simple, and can be implemented by modifying existing *ab initio* packages; (4) it can be parallelized easily, and can scale to thousands of processors; (5) it can be applied to quantum mechanical methods other than the density functional theory.

*Present address: National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

[1] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[2] F. Gygi, R. K. Yates, J. Lorenz, E. W. Draeger, F. Franchetti, C. W. Ueberhuber, B. R. de Supinski, S. Kral, J. A. Gunnels, and J. C. Sexton, Proceedings of the 2005 ACM/IEEE Conference on Supercomputing, 2005 (unpublished).

[3] G. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[4] J.-L. Fattebert and F. Gygi, Phys. Rev. B **73**, 115124 (2006).

[5] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).

[6] It is possible to make the $2 \times 2 \times 2$ fragment smaller (or larger) than $2a$ (a is the size of $m_1 \times m_2 \times m_3$ grid), and then the smaller fragments are defined by the overlaping regions of larger fragments.

[7] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).

[8] F. Shimojo, R. K. Kalia, A. Nakano, and P. Vashishta, Comput. Phys. Commun. **167**, 151 (2005).

[9] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[10] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[11] J. Li and L. W. Wang, Phys. Rev. B **72**, 125325 (2005).

[12] http://hpcrd.lbl.gov/linwang/PEtot/PEtot.html

[13] J. Chelikowsky, Bull. Am. Phys. Soc. **51**, No. 1, Part. 1, 612 (2006).

[14] S. A. Blanton, R. L. Leheny, M. A. Hines, and P. Guyot-Sionnest, Phys. Rev. Lett. **79**, 865 (1997).

[15] M. Shim and P. Guyot-Sionnest, J. Chem. Phys. **111**, 6955 (1999).

[16] L. S. Li and A. P. Alivisatos, Phys. Rev. Lett. **90**, 097402 (2003).

[17] M. Posternak, A. Baldereschi, A. Catellani, and R. Resta, Phys. Rev. Lett. **64**, 1777 (1990).

[18] L. W. Wang and A. Zunger, J. Chem. Phys. **100**, 2394 (1994).

[19] M. E. Schmidt, S. A. Blanton, M. A. Hines, and P. Guyot-Sionnest, J. Chem. Phys. **106**, 5254 (1997).