Two-dimensional limit of exchange-correlation energy functional approximations

Yong-Hoon Kim
Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

In-Ho Lee,* Satyadev Nagaraja, and Jean-Pierre Leburton
Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Randolph Q. Hood†
Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, United Kingdom

Richard M. Martin
Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801
(Received 14 September 1999)

We investigate the behavior of three-dimensional (3D) exchange-correlation energy functional approximations of density-functional theory in anisotropic systems with two-dimensional (2D) character. Using two simple models, the quasi-2D electron gas and two-electron quantum dot, we show a *fundamental limitation* of the local density approximation (LDA) and its semilocal extensions, generalized gradient approximation (GGA) and meta-GGA (MGGA), the most widely used forms of which are worse than the LDA in the strong 2D limit. The origin of these shortcomings is in the inability of the local (LDA) and semilocal (GGA/MGGA) approximations to describe systems with 2D character in which the nature of the exchange-correlation hole is very nonlocal. Nonlocal functionals provide an alternative approach, and explicitly the average density approximation is shown to be remarkably accurate for the quasi-2D electron gas system. Our study is not only relevant for understanding of the functionals but also practical applications to semiconductor quantum structures and materials such as graphite and metal surfaces. We also comment on the implication of our findings to the practical device simulations based on the (semi)local density-functional method.

I. INTRODUCTION

In the Kohn-Sham (KS) density-functional theory[1,2] (DFT), significant efforts have been devoted to improve the local density approximation (LDA).[2] One approach, the generalized gradient approximation,[3–7] (GGA) has been successively improved for the last two decades and now is approaching chemical accuracy (atomization energy errors of order 1 kcal/mol$=0.0434$ eV) with further refinements in the so-called meta-GGA (MGGA).[8] The (M)GGA is desirable in that it leads to better physical quantities for various systems of interest, while it is still computationally cheap due to its semilocal nature. It is clear, however, that any local or semilocal approximation cannot fully reproduce the behavior of the exact nonlocal exchange-correlation energy functional, so one needs to be aware of the limitations of these approximation schemes and the situations where they can break down.

In this paper, we discuss one situation where the local and semilocal approximations of the exchange-correlation energy functional inherently break down: systems with two-dimensional (2D) characteristics, which is relevant to DFT computations of semiconductor devices or other physical systems with 2D character. The original motivation of the current work is recent developments in semiconductor nanotechnology that have achieved quantum dots, which offer enormous technological prospects and allow the study of novel physical phenomena due to dimensionality and electronic correlation effects.[9] Quantum dots can be achieved by confining a 2D electron gas with patterned gates. Semiconductor quantum devices in general involve large ranges of electron densities and density gradients, and the effect of electron-electron interactions can be important, so they provide ideal test cases of the approximate functionals commonly used in DFT. However, although DFT has been already extensively applied to the study of these systems,[10,11] the validity of conventional approximation schemes in these systems has not been fully addressed. Hence, we investigate the robustness of various density-based three-dimensional (3D) local and semilocal exchange and correlation energy-functional approximations, LDA, GGA, and MGGA, in the 2D limit using the idealized quasi-2D electron gas and quantum dot systems. We show that there are inherent limitations resulting from the local or semilocal nature of the exchange-correlation hole in these approximations. Especially, we point out that within the restricted form of the GGA it is very difficult to incorporate the necessary requirement for the 2D limit while at the same time maintaining desirable features of present functionals. We contrast the limitation of these local and semilocal approximations with the nonlocal average density approximation (ADA), and explicitly show the improvement by employing the ADA for the quasi-2D electron-gas system.

The organization of the paper is as follows. In Sec. II, we review the features of the LDA, GGA, MGGA, and ADA necessary for our later discussions. In particular, we emphasize that the approximations in these functionals are essentially approximations to the exchange-correlation hole. In

Sec. III, we first establish the limitation of the local and semilocal approximations by considering the nature of the approximations to the exchange-correlation holes in the 2D limit, and contrast them with the nonlocal approximation (Sec. III A). In Sec. III B we explicitly show this in the 2D homogeneous electron gas with finite thickness: compared with the exact exchange energy which is finite in the 2D limit, the 3D LDA, GGA, and MGGA exchange energies incorrectly diverge to negative infinity. Especially, we point out that the direction of the GGA and MGGA correction to the LDA should be opposite to that of the current forms. This is contrasted with the nonlocal ADA approximation that correctly has a finite 2D limit. In Sec. III C, we investigate an idealized quantum-dot system. By varying the confinement strength along one direction, the system changes its character from 3D to 2D. We show that, while the LDA, GGA, and MGGA give satisfactory descriptions of the isotropic limit, they again fail in the 2D limit. Present (M)GGA's are better than the LDA in the isotropic 3D limit, but they are again worse than the LDA in the 2D limit. In addition, we comment on the validity of 2D and 3D DFT calculations of quantum dots at the experimentally realistic range of anisotropy. In Sec. III D, we discuss density-functional calculations of two physical systems with 2D characters, jellium surface and the graphite. We conclude this paper by summarizing the current work in Sec. IV.

II. EXCHANGE-CORRELATION ENERGY FUNCTIONALS

The exchange-correlation energy may be written as the interaction energy between the electron density $n(\mathbf{r}) = \Sigma_{\sigma=\uparrow,\downarrow} n_\sigma(\mathbf{r})$ and the coupling-constant integrated exchange-correlation hole[12,13] $\bar{\rho}_{xc}([\{n_\sigma\}];\mathbf{r},\mathbf{r}')(\{n_\sigma\} \equiv \{n_\uparrow, n_\downarrow\})$:

$$E_{xc}[\{n_\sigma\}] = \frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \frac{n(\mathbf{r})\bar{\rho}_{xc}([\{n_\sigma\}];\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \quad (1)$$

$$\bar{\rho}_{xc}([\{n_\sigma\}];\mathbf{r},\mathbf{r}') = n(\mathbf{r}') \int_0^1 d\lambda [g^\lambda([\{n_\sigma\}];\mathbf{r},\mathbf{r}') - 1]$$

$$\equiv n(\mathbf{r}')[\bar{g}([\{n_\sigma\}];\mathbf{r},\mathbf{r}') - 1], \quad (2)$$

where $g^\lambda([\{n_\sigma\}];\mathbf{r},\mathbf{r}')$ is the pair-correlation function. We adopt atomic units throughout the paper with $\hbar = e = m_e = 1$. The exact exchange and correlation holes have several important physical conditions that should be also observed by approximations such as the negativity of the exchange hole[4,6]

$$\bar{\rho}_x([\{n_\sigma\}];\mathbf{r},\mathbf{r}') < 0, \quad (3)$$

and sum rules of the exchange and correlation holes[12–14]

$$\int d\mathbf{r}' \bar{\rho}_x([\{n_\sigma\}];\mathbf{r},\mathbf{r}') = -1,$$

$$\int d\mathbf{r}' \bar{\rho}_c([\{n_\sigma\}];\mathbf{r},\mathbf{r}') = 0. \quad (4)$$

Various approximation schemes based on density-derived variables are attempts to approximate the exchange-correlation energy functional

$$E_{xc}[\{n_\sigma\}] = \int d\mathbf{r} n(\mathbf{r}) \epsilon_{xc}([\{n_\sigma\}];\mathbf{r}), \quad (5)$$

or exchange-correlation energy density functional

$$\epsilon_{xc}([\{n_\sigma\}];\mathbf{r}) = \frac{1}{2}\int d\mathbf{r}' \frac{\bar{\rho}_{xc}([\{n_\sigma\}];\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \quad (6)$$

in terms of a function of density and/or other density related variables. In the standard LDA, the exchange-correlation energy density is replaced by that of the homogeneous electron gas at each point $\mathbf{r}$,

$$\epsilon_{xc}([\{n_\sigma\}];\mathbf{r}) \approx \epsilon_{xc}^{LDA}(\{n_\sigma(\mathbf{r})\}) = \epsilon_{xc}^{hom}(\{n_\sigma(\mathbf{r})\}), \quad (7)$$

which can be interpreted as an approximation to the exchange-correlation hole[13,14]

$$\bar{\rho}_{xc}^{LDA}(\{n_\sigma(\mathbf{r})\};\mathbf{r},\mathbf{r}') = n(\mathbf{r})[\bar{g}^{hom}(\{n_\sigma(\mathbf{r})\};|\mathbf{r}-\mathbf{r}'|) - 1]. \quad (8)$$

It is important to notice that the local replacement of density prefactor $n(\mathbf{r}')$ by $n(\mathbf{r})$ in Eq. (8) leads to the LDA exchange-correlation hole being spherical and centered on the electron, while the exact one is centered at another point (such as at the nucleus position in an atom or a molecule) and very asymmetric. Thus we might expect the LDA hole to be a reasonable approximation when the exact exchange-correlation hole is close to the electron. However, since only the spherical average of the exchange-correlation hole

$$\bar{\rho}_{xc}^{SA}(\mathbf{r},R) = \frac{1}{4\pi}\int_\Omega d\mathbf{r}' \bar{\rho}_{xc}(\mathbf{r},\mathbf{r}'), \quad \Omega: |\mathbf{r}-\mathbf{r}'| = R \quad (9)$$

influences the exchange-correlation energy,[13,14] the spherically symmetric nature of the LDA hole does not necessarily represent a poor approximation. In addition, it is known that the LDA is a surprisingly robust approximation scheme, which may be undestood from the fact that its exchange-correlation hole satisfies the hole conditions, Eqs. (3) and (4).[13,14]

In the following sections, we examine important features of other approximations, including the GGA, MGGA, and ADA, which are relevant for our discussions in later sections.

A. Generalized gradient approximation

Although the idea of utilizing density gradient information as a way to improve the LDA was proposed in the original papers of Hohenberg, Kohn, and Sham,[1,2] it is only in the last decade or so in which successful GGA functionals have appeared. This suggests that some of the correct physics of the exchange-correlation effects, which were missing in the original naive gradient expansion approximation (GEA), have been incorporated in recent developments of the GGA. Perhaps, the most important step was the recognition by Perdew and co-workers that the exchange-correlation hole in the GEA does not correspond to a physical hole, nor does it satisfy the negativity condition of the exchange hole [Eq.

(3)] and normalization conditions of the exchange and correlation holes [Eq. (4)].[4,6,15] Following their argument, the GGA can be understood as an approximation of the exchange-correlation hole in which the spurious long-range part of the second-order GEA exchange-correlation GEA hole has been cut off in the real space to satisfy the conditions of Eqs. (3) and (4). One should note that the GGA is based on the modification of the LDA exchange-correlation hole, so its hole is local and it tends to be an improvement over the LDA when the LDA is a good first-order approximation.[16]

Different GGA approaches can be compared by writing the GGA exchange-correlation energy density in terms of the reference LDA exchange energy density multiplied by the enhancement factor F_{xc}^{GGA}:[6,7]

$$\epsilon_{xc}([\{n_\sigma\}];\mathbf{r}) \approx \epsilon_{xc}^{GGA}(\{n_\sigma(\mathbf{r}),|\nabla n_\sigma(\mathbf{r})|\}) \equiv \epsilon_x^{LDA}(n(\mathbf{r}))F_{xc}^{GGA}(\{n_\sigma(\mathbf{r}),|\nabla n_\sigma(\mathbf{r})|\}), \tag{10}$$

where $\epsilon_x^{LDA}(n(\mathbf{r}))=-3[3\pi^2 n(\mathbf{r})]^{1/3}/(4\pi)$. F_{xc}^{GGA} can be naturally divided into two parts, exchange F_x^{GGA} and correlation F_c^{GGA}. For exchange, because of the spin-scaling relation

$$E_x[\{n_\sigma\}]=\tfrac{1}{2}E_x[2n_\uparrow]+\tfrac{1}{2}E_x[2n_\downarrow], \tag{11}$$

we only need to consider the spin-unpolarized case $F_x^{GGA}(n,|\nabla n|)$, which in terms of the dimensionless reduced density gradient

$$s=\frac{|\nabla n|}{2(3\pi^2)^{1/3}n^{4/3}} \tag{12}$$

can be expressed as $F_x^{GGA}(s)$. Numerous gradient approximations for the exchange have been proposed, and in this work we consider the three most successful and popular ones by Becke (B88),[5] Perdew and Wang (PW91),[6] and Perdew, Burke, and Enzerhof (PBE).[7] In Fig. 1, we compare the F_x^{GGA}'s of these three approximations. Most other F_x^{GGA}'s fall between the B88 GGA and PBE GGA,[17] so the qualitative results obtained by employing other functionals can be interpolated from the behaviors of the B88 GGA and PBE GGA.

As shown in Fig. 1, one can divide the GGA into two regions, (i) small $s(0\lesssim s\lesssim 3)$ and (ii) large $s(s\gtrsim 3)$ regions. In region (i), which is relevant for most physical applications, different F_x^{GGA}'s have nearly identical shapes, which explains why different GGA's give similar improvements for many conventional systems with small density-gradient contributions.[18] Most importantly, $F_x^{GGA}\geqslant 1$, so all the GGA's leads to exchange energy lower than the LDA. Typically there are more rapidly varying density regions in atoms than condensed system, so this will lead to the lowering of the exchange energy in atoms more than molecules and solids.[19] This results in the reduction of binding energy, correcting the LDA overbinding and improving agreement with experiment, which is one of the most important characteristics of present GGA's.

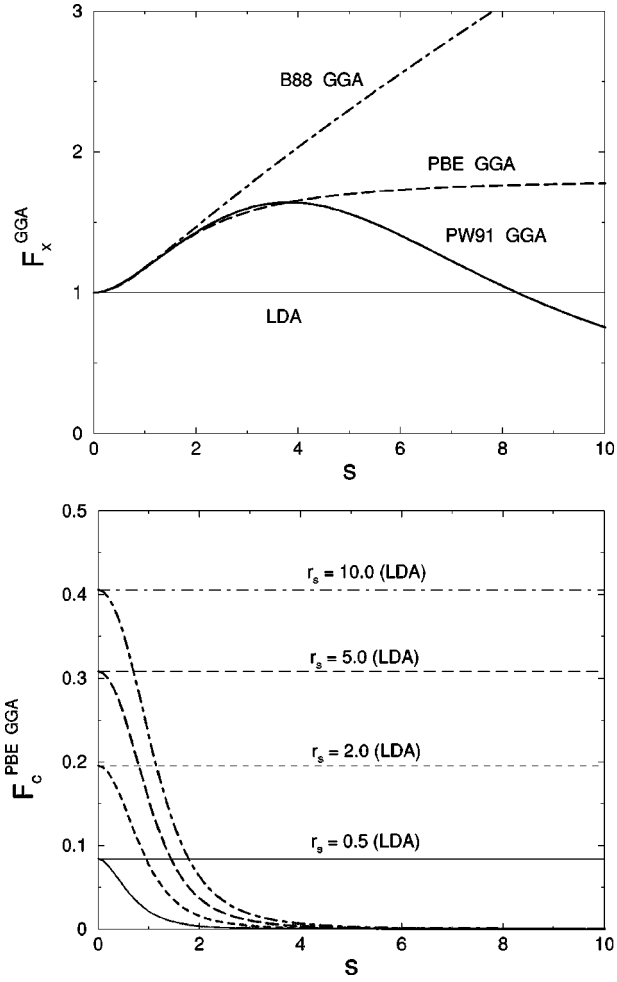

FIG. 1. The enhancement factor over the LDA exchange energy density for (a) B88 GGA, PW91 GGA, and PBE GGA exchange F_x^{GGA} and (b) PBE GGA correlation $F_c^{PBE\ GGA}$ in terms of the dimensionless reduced density gradient $s=|\nabla n|/(2k_F n)$.

In region (ii), different limiting behaviors of F_x^{GGA}'s result from choosing different physical conditions for $s\to\infty$. In the B88 GGA, $F_x^{B88\ GGA}(s)\sim s/\ln(s)$ was chosen to give the correct exchange energy density ($\epsilon_x\to -1/2r$).[5] In the PW91 GGA, choosing $F_x^{PW91\ GGA}(s)\sim s^{-1/2}$ satisfies the Lieb-Oxford bound and the nonuniform scaling condition.[6] In the PBE GGA, the nonuniform scaling condition was dropped in favor of a simplified parametrization with $F_x^{PBE\ GGA}(s)\sim$const.[7] The fact that different physical conditions lead to very different behaviors of F_x^{GGA}'s in region (ii) not only reflects the lack of knowledge of the large-density-gradient regions but also an inherent difficulty of the density-gradient expansion in this region: even if one GGA form somehow gives the correct result for a certain physical property while others fail, it is not guaranteed that the form is superior for other properties in which different physical conditions prevail.

Correlation is more difficult to include, but its contribution to the total energy is typically much smaller than the exchange, especially for systems with large density gradients. Hence, the main qualitative results of this work, which is concerned with the strong anisotropic 2D limit, are determined at the exchange level. For correlations, we employed PW91 GGA and PBE GGA, which are almost identical and designed to be turned off for large-density-gradient regions

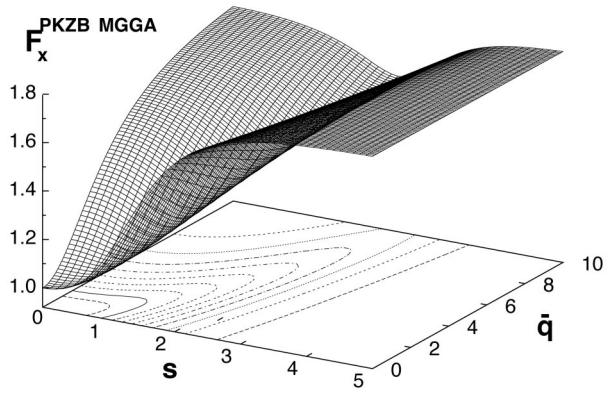

FIG. 2. The exchange part of the PKZB MGGA enhancement factor over the LDA exchange energy density $F_x^{PKZB\ MGGA}$ in terms of the dimensionless reduced density gradient $s=|\nabla n|/(2k_F n)$ and $\bar{q}$ defined in the text.

as shown in Fig. 1(b) for the PBE GGA correlation part enhancement factor $F_c^{PBE\ GGA}$. The fact that correlation decreases with increasing density gradients can be qualitatively understood by the fact that systems with large density gradients have strong confining potentials which increase level spacings and reduce the effect of correlations.

B. Meta-generalized gradient approximation

One natural extension of the GGA is to employ the next higher-order gradient expansion variables, the Laplacian of the density $\nabla^2 n_\sigma(\mathbf{r})$ and the orbital kinetic energy density

$$\tau_\sigma(\mathbf{r})=\frac{1}{2}\sum_{i=1}^{occp.}|\nabla\psi_{\sigma i}(\mathbf{r})|^2, \tag{13}$$

in addition to the density $n_\sigma(\mathbf{r})$ and density gradient $\nabla n_\sigma(\mathbf{r})$. Several forms of this so-called meta-GGA (MGGA) have been suggested, and here we only consider the recent work of Perdew Kurth, Zupan, and Blaha (PKZB) based on the PBE GGA.[8] They removed the dependence on the Laplacian of the density by introducing a new variable

$$\bar{q}=\frac{3\tau}{2(3\pi^2)^{2/3}n^{5/3}}-\frac{9}{20}-\frac{s^2}{12}, \tag{14}$$

which reduces to the dimensionless Laplacian of the density $q=\nabla^2 n/[4(3\pi^2)^{2/3}n^{5/3}]$ in the slowly varying limit but remains finite at a nucleus where q diverges. Hence we can write the PKZB MGGA as

$$\begin{aligned}\epsilon_{xc}([\{n_\sigma\}];\mathbf{r})&\approx\epsilon_{xc}^{MGGA}(\{n_\sigma(\mathbf{r}),|\nabla n_\sigma(\mathbf{r})|,\tau_\sigma\})\\&\equiv\epsilon_x^{LDA}(n(\mathbf{r}))F_{xc}^{MGGA}(\{n_\sigma(\mathbf{r}),|\nabla n_\sigma(\mathbf{r})|,\tau_\sigma\}),\end{aligned} \tag{15}$$

similar to Eq. (10). The enhancement factor $F_x^{PKZB\ MGGA}$ for the exchange is shown in Fig. 2 as a function of s and $\bar{q}$. Unlike the PBE GGA, the PKZB MGGA exchange energy functional satisfies both the correct gradient expansion and the linear response limit for the exchange, and its correlation energy functional is self-interaction-free for a single electron. However, the relevant feature of the PKZB MGGA exchange-correlation energy functional F_x for the current discussion is that it (or PKZB MGGA exchange-correlation hole) is still semilocal and $F_x^{PKZB\ MGGA}$ is always larger than or equal to 1 since it is based on the PBE GGA (see Fig. 2 along the s axis). So the qualitative feature of the PKZB MGGA is similar to that of the PBE GGA, and especially its exchange energy is always lower than the LDA exchange energy.

C. Average density approximation

Two decades ago, Gunnarsson *et al.* criticized the earlier gradient expansion approaches for their failure to satisfy the sum rule [Eq. (4)] and proposed two completely nonlocal approximation schemes, the average density approximation (ADA) and the weighted density approximation (WDA).[14] In the ADA, which has been utilized in this work, the exchange-correlation hole [Eq. (2)] is approximated by

$$\bar{\rho}_{xc}^{ADA}(\{\bar{n}_\sigma(\mathbf{r})\};\mathbf{r},\mathbf{r}')=\bar{n}(\mathbf{r})[\bar{g}^{hom}(\{\bar{n}_\sigma(\mathbf{r})\};|\mathbf{r}-\mathbf{r}'|)-1], \tag{16}$$

leading to

$$\epsilon_{xc}([\{n_\sigma\}];\mathbf{r})\approx\epsilon_{xc}^{ADA}(\{\bar{n}_\sigma(\mathbf{r})\})=\epsilon_{xc}^{hom}(\{\bar{n}_\sigma(\mathbf{r})\}), \tag{17}$$

where

$$\bar{n}(\mathbf{r})=\int d^3\mathbf{r}'\,w(\bar{n}(\mathbf{r});\mathbf{r}-\mathbf{r}')n(\mathbf{r}'), \tag{18}$$

is a nonlocal functional of the density. The important aspect of the ADA is the nonlocal nature of its exchange-correlation hole whose extent does not directly depend upon the density at or around the reference point $\mathbf{r}$ but upon its weighted average. The weight function w could be chosen in several ways. Gunnarsson *et al.* originally proposed a form based on the information of the linear response limit of the homogeneous electron gas. We follow their suggestion and use the weight function w tabulated in their paper,[14] with the Eq. (18) evaluated by the method based on fast Fourier transforms.[20,21]

III. INHERENT LIMITATION OF THE LOCAL AND SEMILOCAL APPROXIMATIONS IN THE ANISOTROPIC 2D LIMIT

A. Basic issues

We first outline the underlying physics by considering the behavior of the exchange-correlation hole of 3D electrons in the 2D limit. The 2D limit of a 3D density can be written as

$$n(\mathbf{r})\rightarrow n^{2D}(\mathbf{r}^{2D})\,\delta(z). \tag{19}$$

If we employ the exact exchange-correlation hole [Eq. (2)], the exchange-correlation energy density [Eq. (6)] in the 2D limit is

$$\begin{aligned}\epsilon_{xc}([\{n_\sigma\}];\mathbf{r})\rightarrow\frac{1}{2}\int d^2\mathbf{r}'^{2D}\\\times\frac{n^{2D}(\mathbf{r}'^{2D})[\bar{g}([\{n_\sigma\}];\mathbf{r}^{2D},\mathbf{r}'^{2D})-1]}{|\mathbf{r}^{2D}-\mathbf{r}'^{2D}|}\end{aligned} \tag{20}$$

which is finite. Note that the Dirac δ function has been removed through the integration along the z direction. On the other hand, when we employ the local LDA or the semilocal GGA or MGGA, density prefactor $n(\mathbf{r}')$ in the exact exchange-correlation hole expression [Eq. (2)] is replaced by $n(\mathbf{r})$. In these cases the Dirac δ function in Eq. (19) will not be removed through the integration in the evaluation of the exchange-correlation energy density as in Eq. (20), which results in the divergence of the exchange-correlation energy density

$$\epsilon_{xc}^{LDA,(M)GGA}(\{n_\sigma(\mathbf{r})\}) \rightarrow -\infty. \tag{21}$$

In conclusion, we can expect the incorrect divergence of the (semi)local approximations due to the approximation of the exchange-correlation hole as being (semi)local.

Now, we contrast this with nonlocal approximations. Specifically, we employ the ADA, which has been described in Sec. II C and will be used in the next subsection. In the ADA, the prefactor $n(\mathbf{r}')$ is replaced by $\overline{n}(\mathbf{r})$ as in Eq. (18), which is finite in the 2D limit:

$$\overline{n}(\mathbf{r}) \rightarrow \overline{n}(\mathbf{r}^{2D}) = \int d^2\mathbf{r}'^{2D} w(\overline{n}(\mathbf{r}^{2D}); \mathbf{r}^{2D} - \mathbf{r}'^{2D}) n(\mathbf{r}'^{2D}). \tag{22}$$

So, we expect the exchange-correlation hole and especially the exchange-correlation energy density in the ADA will correctly have a finite 2D limiting value:

$$\epsilon_{xc}^{ADA}(\{\overline{n}_\sigma(\mathbf{r})\}) \rightarrow \frac{\overline{n}(\mathbf{r}^{2D})}{2} \int d^2\mathbf{r}'^{2D} \times \frac{[\overline{g}^{hom}(\overline{n}_\sigma(\mathbf{r}^{2D}); |\mathbf{r}^{2D} - \mathbf{r}'^{2D}|) - 1]}{|\mathbf{r}^{2D} - \mathbf{r}'^{2D}|}. \tag{23}$$

B. Quasi-2D electron gas

The 2D electron gas is experimentally realizable, for example, in the silicon metal-oxide-semiconductor field-effect transistor (MOSFET) inversion layer and in $GaAs/Al_xGa_{1-x}As$ heterostructures.[22] Although the LDA in the DFT formalism has been typically employed for the study of many-body effect in these device systems since the 1970s,[22] the limitations of the LDA has not been fully resolved.[22,23] In this section we investigate the accuracy of the LDA, GGA, and MGGA for the quasi-2D homogeneous electron gas, which is an idealized model of a quantum well. For carrier with an isotropic effective mass m^*, such as electrons in $GaAs/Al_xGa_{1-x}As$, interacting with Coulomb interactions screened by a dielectric constant ϵ, the Hamiltonian is the same as for electrons in free space if one adopts scaled units of length

$$\tilde{\mathbf{r}} = \alpha \mathbf{r}; \quad \alpha = \frac{m^*}{\epsilon}, \tag{24}$$

and energy

$$\tilde{E} = \beta E; \quad \beta = \frac{\epsilon^2}{m^*}. \tag{25}$$

For GaAs, $m^* = 0.067 m_e$ and $\epsilon = 13.2$, hence the effective unit of energy is $1\ \widetilde{\text{hartree}} = 10.46$ meV and length is $1\ \tilde{a}_0 = 104.22$ Å. However, in the following, we drop the tilde symbol for simplicity, unless explicitly stated otherwise.

A strict 2D electron gas can be characterized by one dimensionless parameter, $r_s^{2D} = 1/\sqrt{\pi n_A}$ or $k_F^{2D} = 2\pi/r_s^{2D}$, where n_A is the areal electron number density, which ranges from 10^{11} to 10^{13} cm^{-2} for typical $GaAs/Al_xGa_{1-x}As$ heterostructures. However, since we are primarily interested in how various 3D DFT exchange-correlation energy approximations perform in the 2D limit, we incorporate the finite thickness by including an envelope wave function $\zeta_0(z)$, with the z direction taken to be perpendicular to the 2D homogeneous electron-gas layer. Here, we assume that only the lowest single 2D subband is populated. Then, within the effective mass approximation, the layer electrons can be characterized, by wave functions of the form

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{A}} \exp(i\mathbf{k}^{2D} \cdot \mathbf{r}^{2D}) \zeta_0(z), \tag{26}$$

where $\mathbf{r}^{2D}$ is the 2D radius vector and $\mathbf{k}^{2D}$ is the 2D wave vector. It is normalized to area A, and $\zeta_0(z)$ is also assumed to be normalized. We take the quantum well potential along the confinement z direction to be parabolic, with the envelope wave function $\zeta_0(z)$ of the form

$$\zeta_0(z) = \left(\frac{b^2}{\pi}\right)^{1/4} e^{-b^2 z^2/2}. \tag{27}$$

In Eq. (27) $1/b$ characterizes the spatial extension of wave functions along the z direction, so we choose the dimensionless ratio b/k_F^{2D} as a measure of the finite thickness effect. Defining the average thickness as $l_0 = \sqrt{2}/b$, which ranges from 20 to 100 Å in experiments, the physically relevant range of $b/k_F^{2D} = r_s^{2D}/l_0$ will be approximately between 1 and 5.

Assuming that the wave function has the form in Eq. (27), Figs. 3(a) and 3(b) show the comparison of the ratios of the exchange and exchange-correlation energy per electron obtained from 3D exact exchange (Hartree-Fock) method and various local/semilocal DFT approximation schemes over the absolute value of the 2D exact exchange energy. The quatntity plotted is the ratio of the energy to the absolute value of the exchange energy in the 2D (large b) limit, $\epsilon_x^{exact,2D} = -(4\sqrt{2})/(3\pi r_s^{2D})$. The finite thickness of the wave function gives the correction $Y(b/k_F^{2D})$

$$\epsilon_x^{exact,3D} = -\frac{4\sqrt{2}}{3\pi r_s^{2D}} Y\left(\frac{b}{k_F^{2D}}\right). \tag{28}$$

The finite thickness correction $Y(b/k_F^{2D})$ in Eq. (28) has been calculated with the envelope-wavefunction Eq. (26) in a similar manner to Ref. 24 where the Fang-Howard envelope-wave function was used.[22] Physically, this finite thickness correction makes the effective interaction softer than the $1/r$ Coulomb interaction for distances small compared to the extension in the z direction of the charge distribution, which leads to a significant correction to the 2D exchange energy for $b/k_F^{2D} \lesssim 10$.

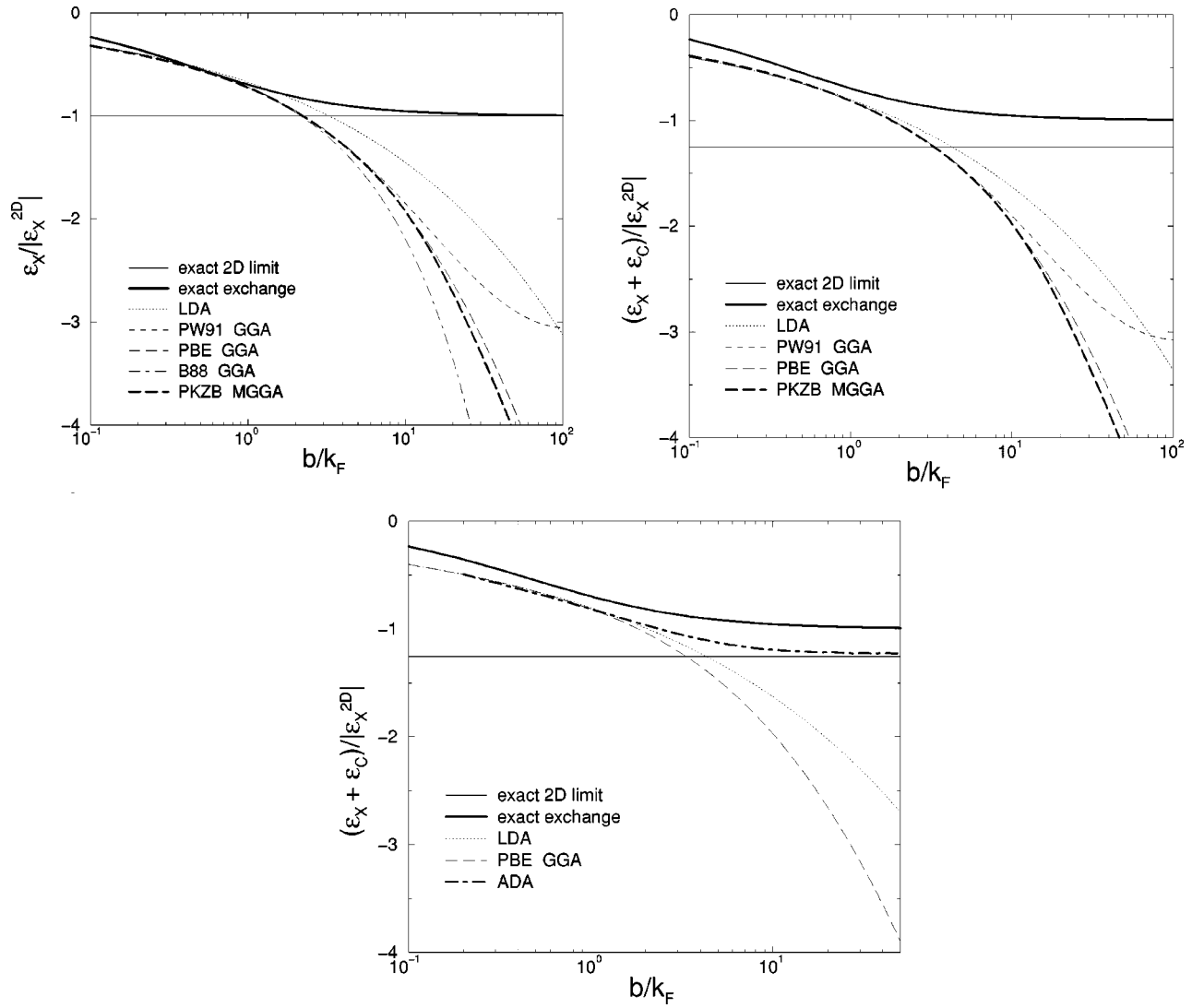

FIG. 3. The ratio of 3D exact, LDA, GGA's and MGGA (a) exchange energy and (b) exchange-correlation energy per electron over the absolute value of the 2D exact exchange energy for the quasi-2D jellium model versus b/k_F^{2D}. (c) The ratio of the 3D ADA exchange-correlation energy per electron over the absolute value of the 2D exact exchange energy for the same system. This ratio is independent of the areal electron number density at the exchange-only level (a), and $r_s = 1.7\tilde{a}_0$ ($n_A = 10^{11}$ cm^{-2}) case has been shown in (b) and (c). The experimentally realistic range of b/k_F^{2D} is about 1–5. For (c), the exact exchange, LDA and PBE-GGA exchange-correlation energies given in (b) are reproduced for comparison. Note that the ADA reduces very closely to the exact 2D exchange-correlation limit.

The very weak confinement regime ($b/k_F^{2D} < 1$) in Fig. 3(a) reveals that the LDA, GGA, and MGGA exchange energies are very close to the exact values. However, as we increase the confinement strength they all go to the wrong limit: the LDA, PBE GGA, B88 GGA, and PKZB MGGA diverge to $-\infty$. Although the PW91 GGA approaches a finite value at large b/k_F^{2D} [due to the fact that $F_x^{PW91\ GGA}(s) \sim s^{-1/2}$ for $s \to \infty$], the magnitude of the converged value is much too large. Note that the LDA is better than the GGA or MGGA for the physically relevant intermediate confinement strength regime, and the direction of the (M)GGA correction should be opposite to that of current forms, i.e., the factor $F_x^{(M)GGA}$ should be less than 1 to reduce the error in the LDA exchange. This feature is closely related with the nonuniform scaling relation, and the PW91 GGA has this property built in, but only for large-s region.[6] In order to describe quasi-2D quantum nanostructures in strong confinement regimes, one needs $F_x^{(M)GGA} \leqslant 1$ for the small s region, but this will apparently worsen the performance of the (M)GGA for other systems such as spherical atoms. This shows that it is difficult use the restricted (M)GGA form to improve both 2D and conventional systems.

The total exchange-correlation energy for $r_s = 1.7\tilde{a}_0$ ($n_A = 10^{11}$ cm^{-2}) is shown in Fig. 3(b), together with the quasi-exact 2D-limit value obtained from the quantum Monte Carlo calculations.[25] The contribution of correlation energy is smaller than the exchange energy in the physically relevant areal density regimes (5–25% for $n_A = 10^{13}$–10^{11} cm^{-2}), so the above qualitative conclusions at the exchange level will not be changed with the inclusion of the correlation energy. One noticeable difference between the LDA and (M)GGA correlation energy functionals is that the magnitude of the LDA correlation energy increases with increasing confinement strength, while that of the (M)GGA decreases due the nature of their correlation form described in Sec. II A.

Now, we explicitly show the different behavior of the

nonlocal approximation that was expected in Sec. III A, by performing the 3D ADA calculations on the quasi-2D electron gas model. The ratio of the ADA exchange-correlation energy and the 2D exact exchange energy is shown in Fig. 3(c) together with the 2D LDA, 3D exact exchange, LDA, and PBE GGA results. The ADA not only correctly reduces to a finite value but also the limit value itself is surprisingly close to the exact 2D limit. This correct limiting behavior of the ADA clearly differentiate it from local or semilocal approximations, and confirms our statements in Sec. III A. Recently, the ADA has been also shown to give improved descriptions of the exchange-correlation energy density over the LDA for the conventional bulk silicon system,[21] but its applications are scarce in the literature due to the difficulty of its implementation.

C. Quantum dot

The second physical system with 2D character we discuss is a GaAs quantum dot. Because the orthogonal dot-growth direction confinement is usually much stronger than the in-plane confinement, the electron distribution in a quantum dot has a pancakelike spheroidal shape. For realistic potential shapes of actual quantum dots, we refer the reader to Figs. 2 and 5 of Ref. 10.

We take a simple model of quantum-dot systems with an anisotropic harmonic oscillator potential as the external potential,

$$V_{ext}(\mathbf{r}) = \tfrac{1}{2}\omega^2(x^2+y^2) + \tfrac{1}{2}\omega_z^2 z^2. \tag{29}$$

as employed in our previous investigation.[11] Here z is the dot-growth direction, and $\omega_z \geqslant \omega$. We consider only two interacting electrons in this potential. The exact exchange energy for this two-electron system is equal to one half of the Hartree energy. In the isotropic potential limit $\omega = \omega_z$, this model can be solved analytically for a discontinuous but infinite set of oscillator frequencies,[26] and a comparative study of the exact KS, LDA, and GGA schemes has been reported recently.[27] These results were used as a check of our calculation method described below.

We performed self-consistent LDA, PBE GGA, and exact exchange[28] (EXX) calculations for $\omega_z/\omega \leqslant 20$, and PW91 GGA and PKZB MGGA energies have been evaluated by the PBE GGA density and wave functions. Technical details of our self-consistent EXX calculations based on the finite-difference grid scheme[11,29] have been presented in Ref. 30. For $\omega_z/\omega > 20$, we used a simple variational EXX approach to generate approximate solutioins: Taking the variational trial wave function as

$$\psi(\mathbf{r}) = \left(\frac{\omega'}{\pi}\right)^{1/2} e^{-\omega'(x^2+y^2)/2} \left(\frac{\omega_z'}{\pi}\right)^{1/4} e^{-\omega_z' z^2/2} \tag{30}$$

with two variational parameters ω' and ω_z' for a given external potential characterized by ω and ω_z, we minimize the EXX total energy of the system and use the variationally optimal wave functions and the corresponding density to evaluate various 3D exchange (and correlation) energies. The self-consistent and variational EXX exchange energies obtained through this procedure show the agreement of 99.8% and 99.4% with the exact KS value for the isotropic case.[27]

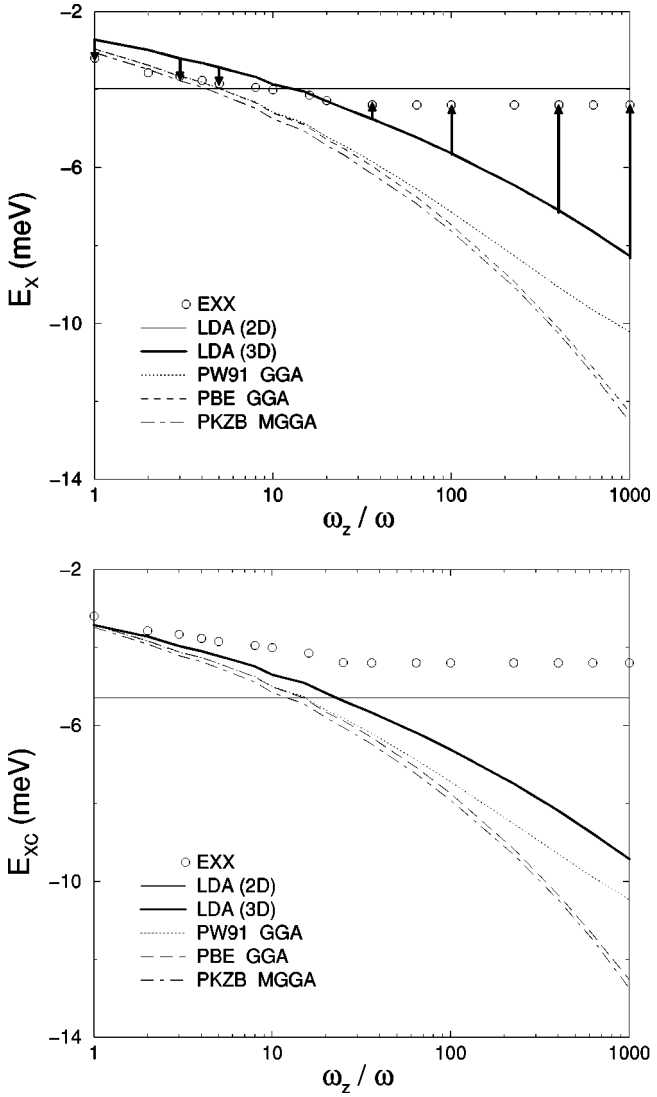

FIG. 4. (a) Exchange energy and (b) exchange-correlation energy of two electrons confined in a quantum dot modeled by an anisotropic harmonic oscillator potential as functions of ω_z/ω. For a fixed lateral-direction external potential $\omega = 2$ meV, the dot-growth direction potential ω_z has been varied from 2 to 2000 meV. The realistic value of ω_z is about 20–50 meV, or $\omega_z/\omega \approx 10$–25. The upward/downward arrows in (a) indicate the direction of correction to the LDA exchange energy. Note that (M)GGA correction is always downward, hence improves the LDA at the isotropic limit, but worsens at the anisotropic limit.

In addition to the 3D DFT calculations, we also performed 2D LDA calculations[25] with the 2D density

$$n^{2D}(\mathbf{r}^{2D}) = 2\left(\frac{\omega''}{\pi}\right) e^{-\omega''(x^2+y^2)}, \tag{31}$$

which has been obtained through a 2D EXX variational minimization procedure as in the large-confinement 3D case ($\omega_z/\omega > 20$) with a single variational parameter ω''.

The exchange and exchange-correlation energies for $\omega = 2$ meV $= 0.1912$ $\widetilde{\text{hartree}}$ are shown in Figs. 4(a) and 4(b) respectively. The EXX exchange values in Fig. 4(a) reveal that the system approach the 2D limit at $\omega_z/\omega \sim 20$–30, and also the 2D LDA exchange value is quite close

to the EXX value in that limit. Now we compare the 3D LDA and (M)GGA exchange energies with the EXX exchange energy. First, at the isotropic limit ($\omega_z/\omega=1$), we can see that the (M)GGA improves the 3D LDA, since the magnitude of the exact exchange energy is larger than that of the 3D LDA exchange energy at $\omega_z/\omega=1$. On the other hand, as we move to the anisotropic limit (large ω_z/ω regime), the 3D LDA and GGA exchange energies become much larger in magnitude than the EXX values, and the GGA's worsen the agreement, as in the 2D electron gas considered in the previous subsection. The required directions of correction to the LDA value have been indicated by upward/downward arrows, which show that we need corrections of opposite signs at the isotropic and very anisotropic limits; however, the (M)GGA correction always has the same sign, which clearly shows the inherent difficulty in constructing (M)GGA's that satisfy both limits.

The qualitative results at the exchange level are not changed with the inclusion of the correlation energy as shown in Fig. 4(b), except that, for small ω_z/ω, the differences between the 3D LDA and (M)GGA exchange-correlation energies are smaller than the exchange case alone, which implies that the 3D LDA is a good approximation with the exchange and correlation together rather than separately.

Before closing this section, we comment on the implication of current findings to experimental situations. Using a value of $\omega_z=20$–50 meV, estimated by the vertical extent of quantum dots (approximately 100 Å), one finds the realistic range of ω_z/ω is about 10–25, which is not the extreme 2D limit where we observed the breakdown of 3D local/semilocal DFT approximations. Hence 3D LDA or (M)GGA quantum-dot simulation results should be reliable for such experimentally realistic problems. Furthermore Fig. 4(b) suggests that 3D and 2D functionals are about equally applicable in this range although 3D functionals are definitely better for $\omega_z/\omega \lesssim 10$.

In addition, by comparing our 3D and 2D EXX models, we can study the effect of employing strictly-2D quantum-dot simulations as have been adopted in many theoretical studies.[9] In Fig. 5, we compare the lateral components of 3D total kinetic energy $\langle K_\uparrow \rangle$, total external potential energy $\langle V_{ext}^{\parallel} \rangle$, and Hartree + exchange energy $\langle V_{ee}^{\parallel} \rangle$, with the corresponding 2D values. Physically, the electron-electron interaction makes the electrons more extended than the noninteracting counterparts, which decreases the kinetic energy and increases the potential energy. On the other hand, the finite thickness of the 2D electron layer has the effect of softening the Coulomb interaction as discussed in Sec. III B, which results in the increase of the kinetic energy and decrease of the potential energy than the 2D limiting values as shown in Fig. 5. This effect is pronounced up to $\omega_z/\omega \sim 20$, and we expect this feature will be missing in the strictly 2D calculations. One more noticeable point is that the anisotropy of the potential induces a bigger change from the isotropic limit in the Hartree + exchange energy (+1.20 meV) than those in the kinetic energy (−0.38 meV) and potential energy (+0.80 meV).

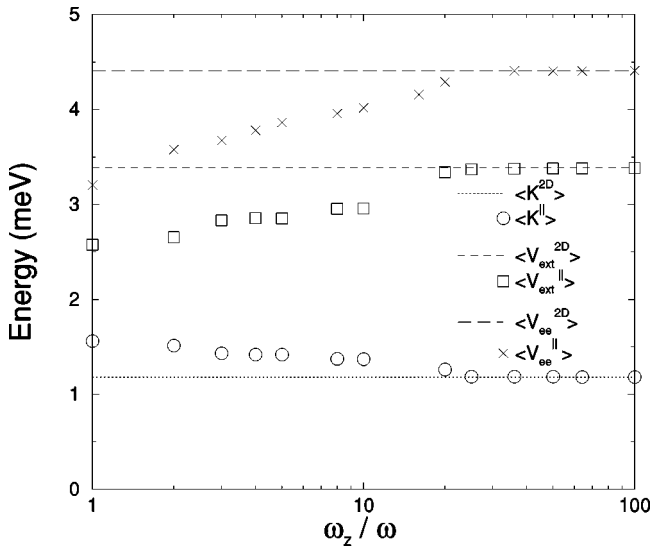

FIG. 5. The lateral components of 3D total kinetic energy $\langle K_\uparrow \rangle$, total external potential energy $\langle V_{ext}^{\parallel} \rangle$, and Hartree + exchange energy $\langle V_{ee}^{\parallel} \rangle$, and the corresponding 2D values $\langle K^{2D} \rangle$, $\langle V_{ext}^{2D} \rangle$, and $\langle V_{ee}^{2D} \rangle$ obtained from variational EXX calculations. For a fixed lateral-direction confinement potential $\omega=2$ meV, the dot-growth direction potential has been varied from 2 to 200 meV.

D. Other systems

In this subsection, we consider two physical systems with 2D character to which our study might have relevance. First is jellium surfaces, which have recently drawn renewed interest due to the discrepancies between different schemes.[31–33] Although we believe quantum Monte Carlo calculation results[32] are the most accurate, it is not our intention here to address this question. Rather, the relevant point we would like to emphasize is that the conclusion of the present study is consistent with that of Rasolt and Geldart on the gradient corrections in the jellium surface:[34] to support their gradient correction coefficient having a different sign from that found in atoms, they emphasized the difference between the localized and extended system and argued that their form is the proper one to use especially for jellium surfaces.[35] Actually, they further brought caution on forcing a ''universal'' gradient approximation form where none exists, which is a warning we feel has been largely ignored. We believe our study supports the argument of Rasolt and Geldart, and especially our second model system, the quantum dot, is a dramatic example showing the nonexistence of a universal GGA form.

The second system we examine is graphite. Graphite and intercalated graphite constitute another large family of materials with quasi-2D electronic properties. Interestingly, in our previous invetigations, we observed that the LDA gives a better description of the energy difference between the diamond and hexagonal graphite structures of C than the PW91 GGA or PBE GGA both at the theoretical and experimental lattice constants.[36] Since the LDA and GGA descriptions of the diamond C are both satisfactory, we can conclude that the GGA descriptions of the graphite is the origin of the problem. To trace the specific source of the error, we decompose the self-consistent total energies at the experimental lattice constants from the LDA, PW91 GGA, and PBE GGA calculations into kinetic, potential, and exchange-correlation energies. The kinetic + potential energy differences between the diamond and graphite structures from the LDA/PBE-GGA calculations are −820/−865 meV, while the

exchange-correlation energy differences between the two structures are 231/386 meV. The fact that the magnitude of the kinetic + potential energy is more than two times bigger than that of the exchange-correlation energy, while the energy difference in the exchange-correlation part is three times bigger than that in the kinetic + potential part, indicates that the description of the graphite in the PBE GGA functional is the main source of the deficiencies. We argue that this can be understood by our findings in the present work. In hexagonal graphite, there are eight electrons per unit cell, and two of them in the π-like state can be considered as 2D electrons, which corresponds to $r_s = 3.54\tilde{a}_0$ or $k_F^{2D} = 0.41\tilde{a}_0^{-1}$. The valence-electron charge-density distribution is known experimentally,[37] so one can estimate the thickness of the electron layer l_0 as $\approx 2.33\tilde{a}_0$ or $b = \sqrt{2}/l_0 = 0.61\tilde{a}_0^{-1}$. This correspond to $b/k_F^{2D} = 1.5$ in Fig. 3, suggesting that the GGA's worse description of the energy difference between the hexagonal graphite and diamond can be explained by our findings in this work that the GGA gives poorer descriptions of 2D systems than the LDA.

IV. SUMMARY AND DISCUSSIONS

The purpose of this work was to show the fundamental limitation of the 3D local/semilocal exchange-correlation energy functional approximations of DFT by considering systems with 2D characteristics. We traced the source of the failure of the LDA and (M)GGA's in 2D systems to the local and semilocal nature of their approximate exchange-correlation holes. These local/semilocal approximations have been contrasted with nonlocal approximations such as the ADA, and the difference has been explicitly demonstrated in our first model, quasi-2D electron gas, in which the LDA and (M)GGA's significantly overestimate while the ADA reproduces the correct limiting behavior. The 2D limit can be considered as a constraint on approximate functionals. This condition is not built in most of the (M)GGA's, and we emphasized that their forms are inherently too restricted to incorporate this requirement while keeping the necessary property to improve the LDA for other conventional systems. Our second example, an anisotropic quantum dot in which we need different signs of the (M)GGA corrections to the LDA at 3D and 2D limits, explicitly shows the danger of expecting a universal gradient approximation form, as pointed out by Rasolt and Geldart.[34] Two other realistic systems, jellium surface and graphite, have been discussed as relevant examples. For practical device simulations, however, we concluded that the LDA and (M)GGA results should be qualitatively correct, as long as experimentally realistic situations are considered.

Note added. Recently, we were informed that Pollack and Perdew have also studied the quasi-2D electron gas results using a different quantum well model, finding results in agreement with ours presented in Sec. III B.[38] They have employed a scaling argument more extensively and also made a connection to the liquid drop model.

ACKNOWLEDGMENTS

We thank Professor K. Burke for suggestion to refer to Ref. 19, Professor J. Perdew for valuable discussion and sharing his work prior to publications, and Dr. V. Rao and Dr. J. Shamaway for suggestions and performing some test calculations. This work was supported by the National Science Foundation under Grants Nos. DMR94-22496 and DMR98-0273.

*Present address: School of Physics, Korea Institute for Advanced Study, Cheongryangri-dong, Dongdaemun-gu, Seoul 130-012, Korea.

†Present address: Lawrence Livermore National Laboratory, Livermore, CA 94550.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).

[3] D.C. Langreth and M.J. Mehl, Phys. Rev. Lett. **47**, 446 (1981); Phys. Rev. B **28**, 1809 (1983).

[4] J.P. Perdew, Phys. Rev. Lett. **55**, 1665 (1985); J.P. Perdew and Y. Wang, Phys. Rev. B **33**, 8800 (1986).

[5] A.D. Becke, Phys. Rev. A **38**, 3098 (1988).

[6] K. Burke, J. P. Perdew, and Y. Wang, in *Electronic Density Functional Theory: Recent Progress and New Directions*, edited by J. F. Dobson, G. Vignale, and M. P. Das (Plenum, New York, 1998).

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

[8] J.P. Perdew, S. Kurth, A. Zupan, and P. Blaha, Phys. Rev. Lett. **82**, 2544 (1999).

[9] L. P. Kouwenhoven, C. M. Marcus, P. L. McEuen, S. Tarucha, R. M. Westervelt, and N. S. Wingreen, in *Mesoscopic Electron Transport*, edited by L. L. Sohn, L. P. Kouwenhoven, and G. Schön (Kluwer Academic, Dordrecht, 1997).

[10] S. Nagaraja, P. Matagne, V.-Y. Thean, J.-P. Leburton, Y.-H. Kim, and R.M. Martin, Phys. Rev. B **56**, 15 752 (1997).

[11] I.-H. Lee, V. Rao, R.M. Martin, J.-P. Leburton, Phys. Rev. B **57**, 9035 (1998); I.-H. Lee, K.-H. Ahn, Y.-H. Kim, and R. M. Martin, Phys. Rev. B **60**, 13 720 (1999).

[12] D.C. Langreth and J.P. Perdew, Solid State Commun. **17**, 1425 (1975).

[13] O. Gunnarsson and B.I. Lundqvist, Phys. Rev. B **13**, 4274 (1976).

[14] O. Gunnarsson, M. Jonson, and B.I. Lundqvist, Phys. Rev. B **20**, 3136 (1979).

[15] J.P. Perdew, K. Burke, and Y. Wang, Phys. Rev. B **54**, 16 533 (1996).

[16] J. P. Perdew (private communication).

[17] J.P. Perdew and K. Burke, Int. J. Quantum Chem. **57**, 309 (1996).

[18] While these three GGA functionals have similar shapes in region (i) the detailed behaviors are different as $s \to 0$. In the B88 GGA (Ref. 5), parameters were chosen empirically resulting in $F_x^{B88\ GGA}(s) \to 1 + 0.235s^2$, while in the PW91 GGA (Ref. 6), the $s \to 0$ limit was chosen to reproduce the correct gradient expansion for small s: $F_x^{PW91\ GGA}(s) \to 1 + 0.123s^2$. In the PBE GGA (Ref. 7), the correct descriptions of the linear response of a uniform electron gas was incorporated in preference to the correct gradient expansion condition resulting in $F_x^{PBE\ GGA}(s) \to 1 + 0.220s^2$.

[19] A. Zupan, K. Burke, M. Ernzerhof, and J.P. Perdew, J. Chem. Phys. **106**, 10 184 (1997).

[20] D.J. Singh, Phys. Rev. B **48**, 14 099 (1993).
[21] R.Q. Hood, M. Y. Chou, A. J. Williamson, G. Rajagopal, R. J. Needs, and M. J. Foulkes, Phys. Rev. Lett. **78**, 3350 (1997).
[22] T. Ando, A.B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[23] F. Stern, Phys. Rev. B **30**, 840 (1984).
[24] F. Stern, Jpn. J. Appl. Phys., Suppl. **2**, Pt. 2, 323 (1974).
[25] Y. Kwon, D.M. Ceperley, and R.M. Martin, Phys. Rev. B **48**, 12 037 (1993).
[26] M. Taut, Phys. Rev. A **48**, 3561 (1993).
[27] C. Filippi, C.J. Umrigar, and M. Taut, J. Chem. Phys. **100**, 1290 (1994).
[28] For the ground-state properties of the example, in which only the lowest single state is occupied, the ''exact exchange'' (EXX) method in DFT is identical to the Hartree-Fock scheme. Discussion on the physical differences between the EXX and Hartree-Fock methods and EXX calculation results of molecular systems are presented in Ref. 30.
[29] Y.-H. Kim, I.-H. Lee, and R. M. Martin, in *Stochastic Dynamics and Pattern Formation in Biological and Complex Systems*, edited by S. Kim, K. Lee, T. K. Kim, and W. Sung (AIP, Woodbury, 1999).
[30] Y.-H. Kim, M. Städele, and R.M. Martin, Phys. Rev. A **60**, 3633 (1999).
[31] E. Krotscheck and W. Kohn, Phys. Rev. Lett. **57**, 862 (1986).
[32] P.H. Acioli and D.M. Ceperley, Phys. Rev. B **54**, 17 199 (1996).
[33] J.M. Pitarke and A.G. Eguiluz, Phys. Rev. B **57**, 6329 (1998).
[34] M. Rasolt and D.J.W. Geldart, Phys. Rev. Lett. **35**, 1234 (1975); D.J.W. Geldart and M. Rasolt, Phys. Rev. B **13**, 1477 (1976).
[35] M. Rasolt and D.J.W. Geldart, Phys. Rev. B **34**, 1325 (1986); Phys. Rev. Lett. **60**, 1983 (1988).
[36] I.-H. Lee and R.M. Martin, Phys. Rev. B **56**, 7197 (1997).
[37] R. Chen, P. Trucano, and R.F. Stewart, Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. **33**, 823 (1977).
[38] L. Pollack and J. P. Perdew, J. Phys.: Condens. Matter (to be published).