# A SUB-LINEAR SCALING ALGORITHM FOR COMPUTING THE ELECTRONIC STRUCTURE OF MATERIALS[*]

## CARLOS J. GARCÍA-CERVERA[†], JIANFENG LU[‡], AND WEINAN E[§]

**Abstract.** We introduce a class of sub-linear scaling algorithms for analyzing the electronic structure of crystalline solids with isolated defects. We divide the localized orbitals of the electrons into two sets: one set associated with the atoms in the region where the deformation of the material is smooth (smooth region), and the other set associated with the atoms around the defects (non-smooth region). The orbitals associated with atoms in the smooth region can be approximated accurately using asymptotic analysis. The results can then be used in the original formulation to find the orbitals in the non-smooth region. For orbital-free density functional theory, one can simply partition the electron density into a sum of the density in the smooth region and a density in the non-smooth region. This kind of partition is not used for Kohn-Sham density functional theory and one uses instead the partition of the set of orbitals. As a byproduct, we develop the necessary real space formulations and we present a formulation of the electronic structure problem for a subsystem, when the electronic structure for the remaining part is known.

**Key words.** sub-linear scaling algorithms, asymptotics, DFT-continuum approximation, density functional theory

**AMS subject classifications.** 35Q40, 74Q05, 34E05

## 1. Introduction

The main purpose of this paper is to introduce a new class of sub-linear scaling algorithms for the analysis of the electronic structure of materials using density functional theory. By sub-linear scaling algorithms, we mean algorithms whose computational cost scales sub-linearly with the size of the system, here the total number of atoms. As was remarked in [6], from an algorithmic viewpoint, one main purpose of multiscale modeling is to develop sub-linear scaling algorithms. In the case of crystalline solids, the well-known quasicontinuum method is an example of a sub-linear scaling algorithm [25, 37], when plastic deformation only occurs on a vanishingly small part of the whole sample. This is the case of interest to us.

Our focus in this paper is electronic structure and density functional theory. We refer to well-known textbooks such as [29, 34] for an introduction to density functional theory, and the following sections for the essential ingredients needed in this article. For now, let us comment briefly on the history of numerical algorithms for density functional theory, particularly the Kohn-Sham model. Early efforts on the numerical solutions of the Kohn-Sham model took the viewpoint that the Kohn-Sham equation is a nonlinear eigenvalue problem and used diagonalization methods to find the eigenvalues and eigenfunctions. The computational complexity of such algorithms scales as $O(N^3)$, where $N$ is the number of plane waves (Fourier modes) used in the discretization of the system. Iterative methods and dynamics-based methods were introduced in the 1980's, sometimes bringing down the cost to $O(N^2)$ or even $O(N \log N)$ [29]. However, these algorithms also require orthogonalization steps, with

[†]Mathematics Department, South Hall, Room 6707, University of California, Santa Barbara, CA 93106, USA (cgarcia@math.ucsb.edu).

[‡]Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA (jianfeng@math.princeton.edu).

[§]Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544, USA (weinan@princeton.edu).

a cost of $O(M_a^3)$ where $M_a$ is the number of atoms in the system. Beginning from the 1990's, much effort has gone into exploring the locality of the problem and developing linear scaling methods [13, 19, 43] with an expected cost of $O(M_a)$. While many interesting ideas have been proposed and explored, a lot remains to be done in order to make these algorithms robust, with a controllable error and with true linear scaling.

The complexity of orbital-based density functional theory has prompted a renewed interest in orbital-free density functional theory [39, 40], and numerical algorithms have been developed for these models [5, 16, 18]. In particular, let us mention the high order accurate, linear scaling algorithms developed in [16], which will be used later in this paper.

At the same time, it has also been realized that a detailed analysis of the electronic structure is only necessary in regions where defects are present or chemical changes are expected. Therefore, beginning in the 1970's, there has been a parallel effort in developing hybrid (or multiscale, multi-physics) algorithms that combine quantum mechanical models with classical mechanical models. The well-known QM-MM algorithm is a good example of this kind of method [14, 42, 45]. These QM-MM algorithms have been quite popular in chemistry. But it is only quite recently that they began to be explored in the context of material modeling [5, 28, 44]. Parallel to the work presented here, there have also been efforts in developing quasicontinuum methods, originally developed as a multiscale method for the analysis of the mechanical deformation of solids based on classical atomistic models [37] for electronic structure study [17, 31]. So far this work is limited to the Thomas-Fermi-von Weiszäcker model, but there is certainly great interest in extending such methods to Kohn-Sham models. Local quasicontinuum methods have been generalized to orbital-free density functional theory [12].

Despite the long history and the popularity of such hybrid or multiscale methods, the issue of consistency and accuracy has never been satisfactorily resolved. This is particularly a concern at the QM-MM interface or, in quasicontinuum methods, at the local-nonlocal interface. For the former class of methods, errors at the QM-MM interface are manifested in the form of unbalanced charges and destabilized chemical bonds. For quasicontinuum methods, such errors result in the well-known "ghost force" [9, 37]. Partly for this reason, there still does not exist a systematic way of improving the accuracy of these methods. Therefore, while they produce some form of approximate solution, one cannot view these methods as systematic ways of solving the original quantum mechanical problem that can achieve any desired accuracy.

Partly motivated by the desire to understand such coupled multiscale methods, we began in [10] an analysis of the continuum limits of DFT models, as well as the accuracy of coupled DFT-continuum models. It was shown, for example, that by modifying the way that long range interaction between the DFT and continuum regions is modeled, one can eliminate the "ghost-force" at the interface between the DFT and the continuum models.

In this paper, we will present asymptotics-based sub-linear scaling algorithms for density functional theory models, with any desired accuracy. Our fundamental assumption is that the positions of the atoms are distributed smoothly except for small isolated regions that contain defects. We use this smoothness assumption to perform asymptotic analysis of the DFT models, with arbitrarily high order accuracy. In the smooth region, we can then solve the DFT models to any specified accuracy requirement, by solving the asymptotic equations that we have derived. These asymptotic equations can be solved on coarse grids whose grid size depends only on the smooth-

ness of the atomic positions and the error tolerance, not on the atomic spacing. The results can then be used to obtain the electronic structure in the remaining non-smooth region, with the solution in the smooth region acting as the environment.

In the present paper, we will focus on the electronic structure problem only with the atomic positions given and fixed. In this case, one solves first for the smooth region, and then the non-smooth region. In other words, the results in the smooth region do not depend on the results in the non-smooth region. However, the philosophy presented here applies equally well to the case when we are also interested in the analysis of the deformation of the material, or simply relaxation of atomic positions. In that case, it is no longer true that the smooth region can be completely solved beforehand, as is the case when classical atomistic models are used.

We now turn to the organization of this paper. In the next section, we introduce the density functional theory models that we will use. This is followed by a discussion of the general strategy of the sub-linear scaling algorithm that we will propose. In Sec. 4, we discuss sub-linear scaling algorithms for the Thomas-Fermi- von Weiszäcker model. In Sec. 5, we discuss the Kohn-Sham model. Some conclusions are drawn in Sec. 6.

## 2. Density functional theory

In Kohn-Sham density functional theory, we consider a set of $N$ wavefunctions or orbitals $\{\psi_k\}, k=1,\ldots,N$, where $N$ is half of the total number of (valence) electrons in the system. There is a factor of one half that comes from the Pauli exclusion principle and the fact that we are not taking into account explicitly the spin variable. The energy functional takes the form:

$$I_\varepsilon(\{\psi_k\}) = \frac{1}{2}\sum_k \int_{\mathbb{R}^3} |\nabla\psi_k(y)|^2\,\mathrm{d}y + \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}(\rho)\rho(y)\,\mathrm{d}y$$
$$+ \frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{(\rho-m)(y)(\rho-m)(y')}{|y-y'|}\,\mathrm{d}y\,\mathrm{d}y', \tag{2.1}$$

where $\rho$ is the electron density:

$$\rho(y) = 2\sum_k |\psi_k(y)|^2. \tag{2.2}$$

In this form, $\{\psi_k(y)\}$ is assumed to satisfy the orthonormality condition,

$$\int_{\mathbb{R}^3} \psi_j^*(y)\psi_k(y)\,\mathrm{d}y = \delta_{jk}. \tag{2.3}$$

The first term in (2.1) is the kinetic energy of the electrons. The third term represents the Coulomb interaction between the electrons. $\epsilon_{\mathrm{xc}}$ is the exchange-correlation functional. This is the term that distinguishes KSDFT from a true first principle such as the quantum many-body problem, and it has to be approximated. For simplicity, we will adopt the local density approximation (LDA) and assume that $\epsilon_{\mathrm{xc}}$ is simply a function of $\rho$: $\epsilon_{\mathrm{xc}} = \epsilon_{\mathrm{xc}}(\rho)$.

The input to the model is the function $m$. Typically $m$ takes the form:

$$m(y) = \sum_i m_i{}^a(y-y_i), \tag{2.4}$$

where the sum is over all the atoms in the system, $y_i$ is the position of the $i$-th atom, and $m_i^a$ is the pseudo-potential associated with the $i$-th atom. The pseudo-potential

tells us the nature of the atom. For simplicity, at this moment, the model only includes local pseudo-potentials. Nonlocal pseudo-potentials [38] will be added in later sections such as Sec. 5.2.

Next we will rescale the model. Let $L_c$ be the characteristic size of the material and $a$ be an atomic length scale such as the lattice constant of a crystal. We rescale the system using $\tilde{x} = x/L_c$, so that after rescaling the system lies in a fixed domain $\Omega$ in $\mathbb{R}^3$. Let $\varepsilon = a/L_c$. We rescale the energy of the system using the Hartree energy constant $E_H = \frac{\hbar^2}{m_e a_0^2} = \frac{e^2}{4\pi\varepsilon_0 a_0}$, where $m_e$, $\hbar$, $e$, $\varepsilon_0$, $a_0$ are electron mass, Planck's constant, unit charge, dielectric constant and Bohr radius respectively. We rescale the quantities as: $\tilde{\rho}(x) = L_c^3 \rho(xL_c)$ and $\tilde{\psi}(x) = L_c^{3/2} \psi(xL_c)$. For example, the rescaled pseudo-potential is given by

$$\tilde{m}^a(y) = m^a(y/\varepsilon)/\varepsilon^3.$$

After this rescaling, the KSDFT functional takes the form:

$$I_\varepsilon(\{\psi_k\}) = \frac{\varepsilon^2}{2} \sum_k \int_{\mathbb{R}^3} |\nabla\psi_k(y)|^2 \, \mathrm{d}y + \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}(\varepsilon^3\rho)\rho(y) \, \mathrm{d}y$$
$$+ \frac{\varepsilon}{2} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{(\rho - m)(y)(\rho - m)(y')}{|y - y'|} \, \mathrm{d}y \, \mathrm{d}y'. \tag{2.5}$$

Notice that after rescaling, both $\epsilon_{\mathrm{xc}}$ and its argument are dimensionless quantities.

A major component of this paper is the asymptotic analysis of the electronic structure problem, using $\epsilon$ as a small parameter. To this end, we need some terminology from continuum mechanics. In continuum theory, the deformation of a solid is described by a displacement field $u : \Omega \to \mathbb{R}^3$ defined on the reference (undeformed or equilibrium) configuration. We assume that the displacement field $u$ has a characteristic length $L_c$. The position of a material point or an atom after deformation is $y = y(x) = x + u(x)$ where $x$ is the position in the reference configuration. $x$ is normally called the Lagrangian coordinate, and $y$, the Eulerian coordinate. One of the problems we face when considering the continuum limit of electronic structure models of solids is the conflict between these coordinates: continuum theory of solids such as elasticity theory naturally uses Lagrangian coordinates and electronic structure theory naturally uses Eulerian coordinates. As a result, we will have to specify carefully the variable with respect to which we carry out differentiation. This inevitably complicates the notation a bit.

We will perform asymptotic analysis in regions where the underlying lattice structure is kept after the deformation, $i.e.,$where there are no defects. Denote by $L$ the underlying crystal lattice. After rescaling, the underlying lattice of the system becomes $\varepsilon L$ with unit cell $\varepsilon\Gamma$. The atoms in the system are located at $x_i \in \varepsilon L \cap \Omega$, $i = 1,\ldots,N$ in equilibrium. We assume that each atom in the system has $n_0$ valence electrons. Therefore, the total number of valence electrons in the system is $n_0|\varepsilon L \cap \Omega|$, where $|\cdot|$ denotes the cardinal number of a set. Since $\{\psi_k\}$ are orthonormal wave functions with total number $n_0|\varepsilon L \cap \Omega|/2$, we have

$$\int_{\mathbb{R}^3} \rho(y) \, \mathrm{d}y = n_0|\varepsilon L \cap \Omega|. \tag{2.6}$$

KSDFT is a rather complicated model. For analytical as well as numerical purposes, it is useful to consider simplified models. One such simplified model is the

Thomas-Fermi-von Weiszäcker (TFW) model. While the Coulomb interaction part is retained, the kinetic part is approximated by functionals that depend only on the density (not on the wave functions):

$$\varepsilon^2 \int_{\mathbb{R}^3} \rho(y)^{5/3}\,\mathrm{d}y + \varepsilon^2 \int_{\mathbb{R}^3} |\nabla\sqrt{\rho(y)}|^2\,\mathrm{d}y, \tag{2.7}$$

where we omit the dimensionless constants such as $C_{TF}, C_{vW}$. For various reasons it is convenient to take $\nu(y) = \sqrt{\rho(y)}$ as the variable for the TFW model. Therefore the energy functional becomes

$$\begin{aligned} I_\varepsilon(\nu) =\,& \varepsilon^2 \int_{\mathbb{R}^3} \nu^{10/3}\,\mathrm{d}y + \varepsilon^2 \int_{\mathbb{R}^3} |\nabla\nu|^2\,\mathrm{d}y + \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}(\varepsilon^3\rho)\rho(y)\,\mathrm{d}y \\ &+ \frac{\varepsilon}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{(\nu^2-m)(y)(\nu^2-m)(y')}{|y-y'|}\,\mathrm{d}y\,\mathrm{d}y'. \end{aligned} \tag{2.8}$$

The TFW functional is minimized under the normalization constraint:

$$\int_{\mathbb{R}^3} \nu(y)^2\,\mathrm{d}y = n_0|\varepsilon L\cap\Omega|.$$

## 3. General strategy for the sub-linear scaling algorithm

Our basic assumption is that, except for isolated defects, the displacement of the atoms follows a smooth vector field, $i.e., u_j = u(x_j)$ for some smooth function $u$ where $u_j$ is the displacement of the $j$-th atom.

The basic intuition is as follows. When the displacement field is smooth, the electron density behaves like

$$\rho^\epsilon(y) \sim \rho_0\left(y, \frac{x(y)}{\epsilon}\right) + \rho_1\left(y, \frac{x(y)}{\epsilon}\right) + \rho_2\left(y, \frac{x(y)}{\epsilon}\right) + \ldots, \tag{3.1}$$

where $y = x + u(x)$, $x(y)$ is the inverse of the map $x \to y = x + u(x)$, and $\rho_0, \rho_1, \ldots$ are smooth functions which are periodic in the second variable with period $\Gamma$. (Recall that $\epsilon$ is the lattice constant and $\Gamma$ is the unit cell of the undeformed lattice). Furthermore, we can derive effective equations for the functions $\rho_0, \rho_1, \ldots$. These effective equations depend only the local behavior of $u$, $i.e.,$derivatives of $u$ at the point of interest. They can be solved independently, as long as we know the derivatives of $u$.

This motivates the following numerical strategy. We first solve these effective equations in the smooth region to find an approximation of the electron density in this region. We then substitute the results in the smooth region into the original problem and solve for the electron density in the non-smooth region.

For KS-DFT, we not only solve for the electron density in the smooth region, we also find a set of localized orbitals. When dealing with the non-smooth region, this set of orbitals defines an environment for the orbitals in the non-smooth region. If we were to use the orthogonal formulation of KS-DFT, we would have to ensure that the orbitals obtained for the non-smooth region are orthogonal to the environment orbitals. This is a difficult constraint to enforce. Therefore we will use a non-orthogonal formulation of KS-DFT. In that case, we only require the orbitals to be independent. This is almost automatic, since the orbitals are localized at different locations.

We emphasize that in this algorithm there is no decomposition of energy, electron density or orbitals, even though we talk about decomposition into smooth and non-smooth regions. We only decompose the set of orbitals into two sets. The only approximation made is to use the effective equations derived from asymptotic analysis.

In the present form, this algorithm also has the feature that the results in the smooth region do not depend on the results in the non-smooth region. Whether this is accurate enough depends on the size of the non-smooth region we choose. To reduce the size of the non-smooth region, one might have to modify the algorithm such that the results of the non-smooth region can be fed back to the smooth region. Such algorithmic improvements will be considered in later publications. In addition, this feature is present because we are limiting ourselves in this paper to the electronic structure problem only. If we are interested in relaxation of the atomic positions under stress, then the smooth and non-smooth regions will be coupled in a non-trivial way.

## 4. The Thomas-Fermi-von Weiszäcker Model

To develop a sub-linear scaling algorithm along the lines just described for the TFW model, we need two main ingredients: The first is asymptotic analysis for the smooth region. The second is an algorithm for the TFW model in real space. We will first discuss these two components.

**4.1. Asymptotics in the smooth region.**    The Euler-Lagrange equations associated with the functional (2.8) are (neglecting the exchange-correlation term for the moment since it does not affect the asymptotics)

$$-\varepsilon^2 \Delta\nu + \frac{5}{3}\varepsilon^2 \nu^{7/3} - \phi\nu + \lambda\nu = 0, \tag{4.1}$$

$$-\Delta\phi = 4\pi\varepsilon(m - \nu^2), \tag{4.2}$$

where $\lambda$ is a Lagrange multiplier for the normalization constraint and $\phi$ is the Coulombic potential generated by the electrons and the ions.

There are two scales in this problem: The scale of the displacement field which is $O(1)$, and the scale of the electron density which is $O(\varepsilon)$. To represent these two scales, we take the following ansatz:

$$\nu = \nu\left(y, \frac{x}{\varepsilon}\right) = \frac{1}{\varepsilon^{3/2}}\nu_0\left(y, \frac{x}{\varepsilon}\right) + \frac{1}{\varepsilon^{1/2}}\nu_1\left(y, \frac{x}{\varepsilon}\right) + \varepsilon^{1/2}\nu_2\left(y, \frac{x}{\varepsilon}\right) + \cdots, \tag{4.3}$$

$$\phi = \phi\left(y, \frac{x}{\varepsilon}\right) = \phi_0\left(y, \frac{x}{\varepsilon}\right) + \varepsilon\phi_1\left(y, \frac{x}{\varepsilon}\right) + \varepsilon^2\phi_2\left(y, \frac{x}{\varepsilon}\right) + \cdots, \tag{4.4}$$

$$\lambda = \lambda_0 + \varepsilon\lambda_1 + \varepsilon^2\lambda_2 + \cdots, \tag{4.5}$$

where $x = x(y) = (\boldsymbol{I} + u)^{-1}(y)$ is the Euler-Lagrange map, which gives the Lagrangian coordinate corresponding to $y$. $\nu(y, z)$ and $\phi(y, z)$ are functions defined on $\mathbb{R}^3 \times \Gamma$ and are periodic in the second variable on $\Gamma$. Note that we use the Eulerian coordinate for the first variable and the Lagrangian coordinate for the second variable. The reason for this is that we will impose periodicity in the second variable with a *fixed* period which is going to be the unit cell $\Gamma$ (in atomic units) in the equilibrium configuration. Similarly, we have:

$$m(y) = m\left(y, \frac{x}{\varepsilon}\right) = \frac{1}{\varepsilon^3}m_0\left(y, \frac{x}{\varepsilon}\right) + \frac{1}{\varepsilon^2}m_1\left(y, \frac{x}{\varepsilon}\right) + \frac{1}{\varepsilon}m_2\left(y, \frac{x}{\varepsilon}\right) + \cdots. \tag{4.6}$$

For example, from (2.4), one can see that

$$m_0(y, z) = \sum_{z_i \in L} m_0^a((\boldsymbol{I} + \boldsymbol{A})(x)(z - z_i)), \tag{4.7}$$

where $\boldsymbol{A} = \nabla u$ is evaluated at $x = x(y)$.

We will use the identity

$$-\Delta f\left(y,\frac{x}{\varepsilon}\right)=-\Delta_1 f\left(y,\frac{x}{\varepsilon}\right)-\frac{1}{\varepsilon}\nabla_2^x\cdot\nabla_1 f\left(y,\frac{x}{\varepsilon}\right)-\frac{1}{\varepsilon}\nabla_1\cdot\nabla_2^x f\left(y,\frac{x}{\varepsilon}\right)-\frac{1}{\varepsilon^2}\Delta_2^x f\left(y,\frac{x}{\varepsilon}\right).$$

Here $\nabla_1$ denotes gradient in the first variable, and $\nabla_2$, the second. We write $\nabla_2^x=(\boldsymbol{I}+\nabla u(x))^{-\mathrm{T}}\nabla_2$ and $\Delta_2^x=\nabla_2^x\cdot\nabla_2^x$ to simplify the notation. The factors $(\boldsymbol{I}+\nabla u(x))^{-\mathrm{T}}$ arise as a result of the transformation between Eulerian and Lagrangian coordinates.

Expanding the Euler-Lagrange equations, we get the leading order equations

$$-\Delta_2^x\nu_0+\frac{5}{3}\nu_0^{7/3}-\phi_0\nu_0+\lambda_0\nu_0=0, \tag{4.8}$$

$$-\Delta_2^x\phi_0=4\pi(m_0-\nu_0^2). \tag{4.9}$$

The next order equations are

$$-\Delta_2^x\nu_1-\nabla_1\cdot\nabla_2^x\nu_0-\nabla_2^x\cdot\nabla_1\nu_0$$
$$+\frac{35}{9}\nu_0^{4/3}\nu_1-\phi_1\nu_0-\phi_0\nu_1+\lambda_1\nu_0+\lambda_0\nu_1=0, \tag{4.10}$$

$$-\Delta_2^x\phi_1-\nabla_1\cdot\nabla_2^x\phi_0-\nabla_2^x\cdot\nabla_1\phi_0=4\pi(m_1-2\nu_0\nu_1). \tag{4.11}$$

The leading order equations (4.8)–(4.9) are a set of equations for functions defined on the unit cell $\Gamma$, with $\boldsymbol{A}=\nabla u(x)$ as parameters.

This asymptotic procedure bears a lot of similarity with homogenization theory. As was observed in [11], many problems in homogenization theory are best viewed from a fiber bundle viewpoint in which the macro-scale variations play the role of the base manifold and the local micro-structure plays the role of the fibers. For the example discussed here, the base manifold describes the smooth distribution for the positions of the nuclei, and the fibers describe the local electronic structure represented by the periodic problem in the fast variable $z$.

**4.2. Real space formulation and algorithms.**    We now review briefly the general ideas that will be used to solve the TFW model. We will use the algorithm presented in [16], and we refer to that article for details.

The algorithm consists of two main components: the discretization and optimization of the energy functional.

To discuss the discretization of the energy functional, let us focus on the simple case when the domain is given by $D=[0,L_1]\times[0,L_2]\times[0,L_3]$ in the Eulerian coordinates. The domain $D$ is discretized using a uniform mesh with grid sizes $\Delta y_1=L_1/n_1$, $\Delta y_2=L_2/n_2$, and $\Delta y_3=L_3/n_3$. The density is defined at the center of the cells: $\nu_{i,j,k}\approx\nu(y_{1,i},y_{2,j},y_{3,k})$, where $y_{1,i}=(i-\frac{1}{2})\Delta y_1$, $i=0,\ldots,n_1+1$, $y_{2,j}=(j-\frac{1}{2})\Delta y_2$, $j=0,\ldots,n_2+1$, $y_{3,k}=(k-\frac{1}{2})\Delta y_3$, $k=0,\ldots,n_3+1$. The grid points with indices 0 or $n_1+1$, $n_2+1$, and $n_3+1$ are ghost cells outside the computational domain, and are used to impose the boundary conditions. Since the density is known on the boundary, the ghost values are defined using second order extrapolation. On the $y_1=0$ boundary, where the boundary value is $\nu_W$, the ghost values are defined as $\nu_{0jk}=2\nu_{W,jk}-\nu_{1jk}$. A similar expression is used on the other boundaries.

The gradient of $\nu$ is approximated using forward differences. The integral of the gradient term in the energy is approximated using the trapezoidal rule, and the remaining terms are discretized using the midpoint rule. We denote the discretized energy functional by $F_h[\nu]$.

To approximate the convolution with the Hartree potential we approximate $\nu$ by a piecewise constant function:

$$\left(K_H * \nu^2\right)(y_{1,i}, y_{2,j}, y_{3,k}) \approx \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} \sum_{p=1}^{n_3} \nu_{rsp}^2 \int_{D_{rsp}} \frac{1}{|y_{ijk} - y|} \, \mathrm{d}y, \qquad (4.12)$$

where $D_{rps} = [y_{1,r} - \Delta y_1/2, y_{1,r} + \Delta y_1/2] \times [y_{2,s} - \Delta y_2/2, y_{2,s} + \Delta y_2/2] \times [y_{3,p} - \Delta y_3/2, y_{3,p} + \Delta y_3/2]$, and $y_{ijk} = (y_{1,i}, y_{2,j}, y_{3,k})$. The integrals in (4.12) can be evaluated using a combination of explicit integration and adaptive Gaussian quadrature [23], and depend only on the distance between $y_{ijk}$ and $y_{rsp}$. Defining

$$K_{i-r,j-s,k-p} = \int_{D_{rsp}} \frac{1}{|y_{ijk} - y|} \, \mathrm{d}y, \qquad (4.13)$$

we can write (4.12) as a discrete convolution:

$$(K_H * \nu^2)_{ijk} = \sum_{r=1}^{n_x} \sum_{s=1}^{n_y} \sum_{p=1}^{n_z} \nu_{rsp}^2 K_{i-r,j-s,k-p}, \qquad (4.14)$$

which can be evaluated using the Fast Fourier Transform (FFT) [33].

We use a modification of the Truncated Newton method for energy minimization, appropriate for constrained minimization under the constraints $\|\nu\|_2^2 = N$, and $\nu \geq 0$. In Newton-based minimization methods, given an approximation $\nu$, the energy is approximated around $\nu$ by a quadratic functional, which is subsequently minimized to obtain a descent direction $p$:

$$F\left[\frac{\nu + p}{\|\nu + p\|}\right] = F[\nu] + (G[\nu], p)_h + \frac{1}{2}(H[\nu] \cdot p, p)_h + O(|p|^3), \qquad (4.15)$$

where $(\cdot, \cdot)_h$ denotes the discrete inner product, given by

$$(u, v)_h = \Delta y_1 \Delta y_2 \Delta y_3 \sum_{i,j,k} u_{ijk} v_{ijk}. \qquad (4.16)$$

In minimization without constraints, $G$ and $H$ are the gradient and Hessian of the energy, respectively. For the constrained problem, $G$ and $H$ are projected versions of the gradient and Hessian, respectively. The projected gradient is

$$G[\nu] = \nabla F[\nu] - (\nabla F[\nu], \nu)_h \nu, \qquad (4.17)$$

where $\nabla F[\nu]$ is the unconstrained gradient, obtained from the discrete energy by taking partial derivatives with respect to the variables $\{\nu_{ijk}\}$.

Minimization of the quadratic part of (4.15) with respect to $p$ leads to the linear equation

$$H[\nu] \cdot p = -G[\nu]. \qquad (4.18)$$

The matrix $H[\nu]$ is symmetric, but not necessarily positive definite. We use the Preconditioned Conjugate Gradient (PCG) to solve equation (4.18). For the preconditioner we use the Laplacian term, which in our discretization can be inverted using the FFT. If $H[\nu]$ is not positive definite, the procedure fails by producing a direction

of negative curvature. In that case we use the corresponding approximation to the solution of system (4.18) as our descent direction. Sufficiently near the minimum, $H[\nu]$ becomes positive definite, and from (4.18) we get $p = -H[\nu]^{-1} \cdot G[\nu]$, *i.e.,*Newton's method. For details regarding the convergence of this algorithm in the unconstrained case, see [32].

Given the current approximation, $\nu$, and a descent direction, $p$, the next approximation is computed with a line search using

$$f(\epsilon) = F\left[\frac{|\nu + \epsilon p|}{\|\nu + \epsilon p\|}\right]. \tag{4.19}$$

In the line search we impose the Wolfe conditions to ensure sufficient decrease in the energy [32].

It is relatively straightforward to extend this algorithm to account for periodic boundary conditions. The only necessary changes are the discretization and the computation of the Coulomb potential.

In the periodic case, we use a pseudo-spectral, Fourier collocation method [21]. The values of the density are given at the grid points $(y_{1,i}, y_{2,j}, y_{3,k}) = (i\Delta y_1, j\Delta y_2, k\Delta y_3)$, where $i = 0, \ldots, n_1 - 1$, $j = 0, \ldots, n_2 - 1$, and $k = 0, \ldots, n_3 - 1$. Given a local deformation gradient $\boldsymbol{I} + \boldsymbol{A} = \boldsymbol{I} + \nabla u$, the Coulomb interactions are obtained by solving the equation

$$-((\boldsymbol{I} + \boldsymbol{A})^{-\mathrm{T}}\nabla)^2 \phi = 4\pi(\nu^2 - m), \tag{4.20}$$

with periodic boundary conditions. This can be done efficiently using the FFT.

**4.3. Sub-linear scaling algorithm for the TFW model.** The basic assumption is that the distribution (or displacement) of the atoms follows a smooth function except for a very small part of the sample. This allows us to divide the whole sample into smooth and non-smooth regions, denoted by $\Omega_s$ and $\Omega_{ns}$ respectively.

We showed that the Euler-Lagrange equation in the smooth region can be solved asymptotically through a series of much simplified equations. For the electronic structure problem, these asymptotic equations are periodic problems for the deformed unit cells. Moreover, they are independent of the electronic structure in the non-smooth region. As a consequence of the smoothness of the displacement, there is no need to solve the periodic problems for each unit cell. It is enough to consider a subset of the unit cells that form a macro-grid covering the smooth region, and solve the periodic problems on these unit cells, as depicted in Figure 4.1. The electronic structure on other unit cells can be obtained through interpolation. The size of the macro-grid is determined by the smoothness of the displacement field and the error tolerance.

The algorithm proceeds as follows. We first select a subset of unit cells (whose centers lie on the grid points of a macro grid) in the smooth region, and solve the asymptotic equations on these unit cells. We then find the electronic structure in the non-smooth region, $\rho_{ns}$, by minimizing the original energy functional, using the results already obtained for $\rho_s$ in $\Omega_s$. In addition, we impose the boundary condition for $\rho_{ns}$ at the boundary of $\Omega_{ns}$:

$$\rho_{ns} = \rho_s. \tag{4.21}$$

To complete the description of this algorithm, we need to
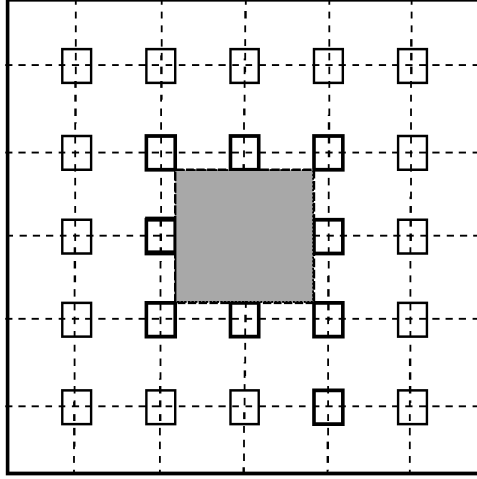  1. discuss how to solve the asymptotic equations in the smooth region,

FIG. 4.1. *Decomposition of the domain into smooth and non-smooth regions. The non-smooth region is shown in dark gray. In the smooth region, a coarse grid is defined. At the grid points in the smooth region, the unit cell is depicted. The asymptotic equations are solved on these unit cells, using the local deformation $\boldsymbol{A} = \nabla u(x)$ as a parameter.*

    2. discuss the interpolation procedure for extending results from a subset of the unit cells to the whole smooth region, and

    3. discuss how to solve for $\rho_{ns}$.

These issues will be discussed in detail now.

    **4.4. Algorithm for the smooth region.**  In the region where the atoms are smoothly distributed, the electronic states can be thought of as a smoothly varying microstructure (see the illustration in figure 4.2). The microscale variation is the variation of the electronic structure inside a unit cell. The macroscale variation is the variation of the smooth displacement field, which also induces a variation of the electronic structure. Furthermore, as we saw earlier, the microstructure is entirely local. To compute such a microstructure, we can proceed using the following steps.

    *Step 1.* We select a coarse grid of unit cells, as discussed earlier.

    *Step 2.* We solve the asymptotic equations (4.8)-(4.9) on these unit cells. To do this, let us define the periodic Thomas-Fermi-von Weiszäcker functional (putting back the exchange-correlation term),

$$I_{\boldsymbol{A}}(\nu) = \int_{\Gamma} \nu^{10/3}(z;\boldsymbol{A})\,\mathrm{d}z + \int_{\Gamma} |(\boldsymbol{I}+\boldsymbol{A})^{-\mathrm{T}}\nabla\nu(z;\boldsymbol{A})|^2\,\mathrm{d}z + \int_{\Gamma} \epsilon_{\mathrm{xc}}(\nu^2)\nu^2(z;\boldsymbol{A})\,\mathrm{d}z$$
$$+ \frac{1}{2}\iint_{\Gamma\times\Gamma}(\nu^2 - m_{\mathrm{CB}})(z;\boldsymbol{A})G(z-z';\boldsymbol{A})(\nu^2 - m_{\mathrm{CB}})(z';\boldsymbol{A})\,\mathrm{d}z\,\mathrm{d}z', \quad (4.22)$$

where $m_{\mathrm{CB}}(z;\boldsymbol{A}(x)) = m_0(y,z) = \sum_{z_i} m_0^a((\boldsymbol{I}+\boldsymbol{A})(x)(z-z_i))$, and $G(z;\boldsymbol{A})$ is the periodic Coulombic kernel [27], which is the solution of

$$-((\boldsymbol{I}+\boldsymbol{A})^{-\mathrm{T}}\nabla)^2 G(z;\boldsymbol{A}) = 4\pi\left(\delta(z) - \frac{1}{|\Gamma|}\right), \quad \text{in } \Gamma, \qquad (4.23)$$

satisfying the periodic boundary condition and

$$\lim_{z\to 0}\big(G(z;\boldsymbol{A})-1/|(\boldsymbol{I}+\boldsymbol{A})z|\big)=0.$$

It is easy to see that the Euler-Lagrange equations of $I_{\boldsymbol{A}}(\nu)$, with the normalization constraint

$$\int_{\Gamma}\nu^2(z;\boldsymbol{A})\,\mathrm{d}z=\frac{n}{\det(\boldsymbol{I}+\boldsymbol{A})},\qquad(4.24)$$

are the same as (4.8)-(4.9) with $\boldsymbol{A}=\nabla u(x)$. Therefore the leading order asymptotic equations can be solved using the minimization algorithm described above. Often the results already give us an accurate approximation of the electronic structure in these regions. We rarely need to solve the higher order equations. If necessary, we can solve the higher order equations using standard methods, since they are linear problems.

Note that in [1], Blanc *et al.* . rigorously proved that the integration of (4.22) does indeed give the leading order energy.

*Step 3.* The electronic structure at other locations is obtained through the following multiscale interpolation procedure.

Assume that we have a function of the form

$$u^\epsilon(x)=U\left(x,\frac{x}{\epsilon}\right),$$

where $U$ is a smooth function and periodic in the second variable, with period $\Gamma$. Consider a macroscale grid with grid points $\{x_j\}$, and let $U_j(z)=U(x_j,z)$. The multiscale interpolation problem we are facing can be formulated as follows: Knowing the functions $\{U_j\}$, and given $x\in\Omega$, how do we approximate the function $U^x(z)=U(x,z)$?

Without the dependence on $z$, this is a standard interpolation problem. With the dependence on $z$, this is still a standard interpolation problem: all we have to do is to view $z$ as a parameter and perform standard interpolation.

Assume that we have a standard interpolation operator in the form

$$(I_h u)(x)=\sum_j a_j(x)u_j,$$

where $\{a_j\}$ is some weight function. We may then simply write

$$(I_h U)(x,z)=\sum_j a_j(x)U_j(z).$$

Going back to the multiscale representation, we have that

$$(I_h u^\epsilon)(x)=\sum_j a_j(x)U_j\left(\frac{x}{\epsilon}\right).$$

For example, in one dimension, for the piecewise linear interpolation, we have

$$(I_h f)(x)=\frac{x_j-x}{\Delta x}f(x_{j-1})+\frac{x-x_{j-1}}{\Delta x}f(x_j)$$

if $x\in(x_{j-1},x_j]$. Then for the multiscale problem, we use

$$(I_h U)(x,z)=\frac{x_j-x}{\Delta x}U(x_{j-1},z)+\frac{x-x_{j-1}}{\Delta x}U(x_j,z).$$

Once we have solved the local problems on a coarse grid, we can then use this interpolation procedure to find the electronic structure everywhere in the smooth region. In particular, we can use this procedure to find the boundary values needed for the non-smooth region, as we explain below.

**4.5. Algorithm for the non-smooth region.**    In the non-smooth region $\Omega_{ns}$, we minimize the energy

$$
I(\nu) = \int_{y(\Omega_{ns})} |\nabla \nu|^2 \, \mathrm{d}y + \int_{y(\Omega_{ns})} \nu^{10/3} \, \mathrm{d}y + \int_{y(\Omega_{ns})} \epsilon_{\mathrm{xc}}(\nu^2)\nu^2 \, \mathrm{d}y
$$
$$
+ \frac{1}{2} \iint_{y(\Omega_{ns}) \times y(\Omega_{ns})} \frac{(\nu^2 - m_{ns})(y)(\nu^2 - m_{ns})(y')}{|y - y'|} \, \mathrm{d}y \, \mathrm{d}y' + \int_{y(\Omega_{ns})} (\nu^2 - m_{ns})V_s(y) \, \mathrm{d}y,
$$

$$(4.25)$$

subject to the constraints $\nu_{ns} \geq 0$ and

$$
\int_{y(\Omega_{ns})} \nu_{ns}^2 \, \mathrm{d}y = N_{ns},
\tag{4.26}
$$

where $N_{ns}$ is the number of electrons in the non-smooth region, and with boundary condition

$$
\nu_{ns}(y) = \nu_s(y), \quad y \in \partial y(\Omega_{ns}).
\tag{4.27}
$$

If necessary, $\nu_s$ is obtained through the multi-scale interpolation procedure described above. Here $V_s$ contains the Coulomb potential due to the electron density in the smooth region,

$$
V_s(y) = \int_{y(\Omega_s)} \frac{(\nu_s^2 - m_s)(y')}{|y - y'|} \, \mathrm{d}y', \quad y \in y(\Omega_{ns}).
\tag{4.28}
$$

Notice that we use $y(\Omega_{ns})$ to denote the non-smooth region after deformation (or in the Eulerian coordinate system). For the exchange and correlation energy in (4.25), we use the local density approximation (LDA) and specifically the Ceperley-Alder functional [4], as parametrized by Perdew and Zunger [35]. The atomic potential $m_{ns}$ is generated by the atoms located inside the non-smooth region. The minimization problem for (4.25) can be solved using the same algorithm as described in Sec. 4.2. The main difficulty is the treatment of the Coulomb potential $V_s$, defined by (4.28). Here we discuss two different ways of treating this Coulomb term, and refer the reader to the article [15], where a more thorough discussion is presented.

The first approach is the fast multipole method (FMM) [22]. FMM is a well-known linear scaling algorithm for computing Coulomb interactions. However, FMM cannot be applied to our problem directly since we have a two-scale charge distribution (see Figure 4.2). To deal with this multiscale issue, we adopt a HMM (heterogeneous multi-scale method [7]) philosophy, namely we apply FMM (which acts as the macro-solver in HMM) on a set of effective charges whose moments are computed from the detailed multiscale charge distribution. For a detailed description of this approach, see [15].

In the absence of large scale elastic deformations, a simpler approach can be obtained by combining a Poisson solver in the smooth region with a convolution in the non-smooth region. Since the lattice is not deformed, the smooth density can be extended periodically to the whole space. Therefore the potential can be written as

$$
V_s(y) = \int_{\mathbb{R}^3} \frac{(\nu_s^2 - m_s)(y')}{|y - y'|} \, \mathrm{d}y' - \int_{y(\Omega_{ns})} \frac{(\nu_s^2 - m_s)(y')}{|y - y'|} \, \mathrm{d}y'.
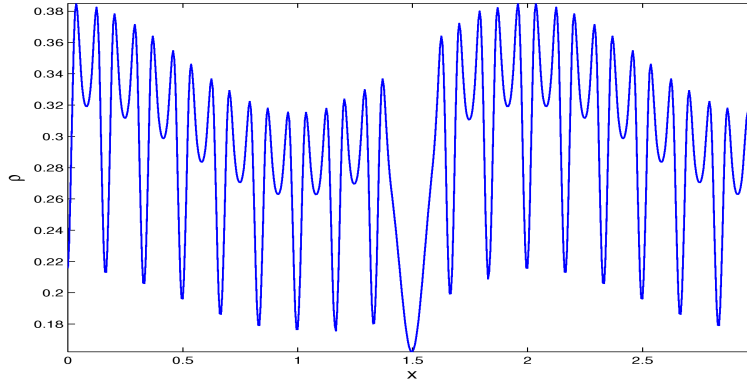\tag{4.29}
$$

FIG. 4.2. *Electronic density around a vacancy in an aluminum FCC crystal. The oscillations at the atomic level are modulated by a smooth deformation.*

The first term can be obtained by solving the equation

$$-\Delta\phi = 4\pi(\nu_s^2 - m_s) \qquad (4.30)$$

in the unit cell, with periodic boundary conditions, and extending $\phi$ periodically to $\mathbb{R}^3$. The second term in (4.29) can be evaluated as a convolution using the FFT, as described in Sec. 4.2.

**4.6. Numerical results: the vacancy problem.**      As an illustration, we have computed the electronic structure of an undeformed aluminum FCC crystal. In our first example, we consider a perfect (infinite) crystal, and in our second example we consider a crystal with a vacancy, where one atom has been removed. Vacancies in a crystal lattice are commonly studied using the so-called *supercell method*. In that approach, a large region containing the vacancy is considered, and the energy functional is minimized using periodic boundary conditions at the boundary of the large region. One drawback of this approach is that the supercell must be exceedingly large in order to reduce the interactions between the artificially created periodic copies of the defect. In the approach presented here, this is not necessary. The domain is divided into two regions: the non-smooth region, $\Omega_{ns}$, which surrounds the vacancy, and the smooth region, which in this case is simply $\mathbb{R}^3 \setminus \Omega_{ns}$.

In this example, the non-smooth region is formed by $4^3 = 64$ unit cells, and the missing atom is located at the center of this domain. In the case of FCC aluminum, each perfect unit cell has 4 atoms, and therefore 12 valence electrons. The periodic cell problem in the smooth region is discretized using $32^3$ grid points, *i.e.,*approximately 4 grid points per atomic unit. The initial density was taken to be random, and normalized to satisfy charge conservation. For the non-smooth region, we require the boundary value at the center of the face adjacent to the smooth region. This value is obtained as the average of the values of $\nu_s$ at the vertices of that face. The external potential is computed as described previously, by extending the periodic potential into the non-smooth region, and subtracting the convolution term. This only needs to be computed once.

The grid size in the non-smooth region is the same as in the smooth region. This is not a limitation, as grid refinement can be performed in the area surrounding the non-smooth region. Although the initial density can be chosen to be a random array
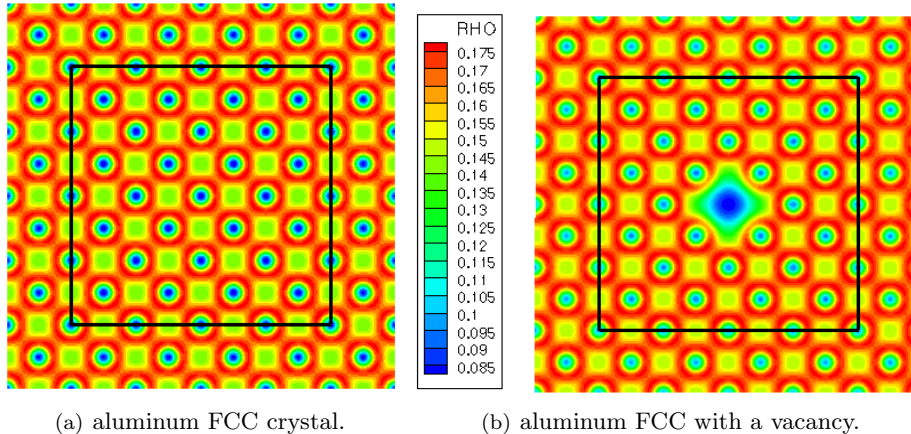
(a) aluminum FCC crystal.          (b) aluminum FCC with a vacancy.

FIG. 4.3. *Contour levels of the electronic density obtained using the Thomas-Fermi-von Weiszäcker model. We only present an interior slice. The box indicates where the separation between the smooth and non-smooth regions is.*
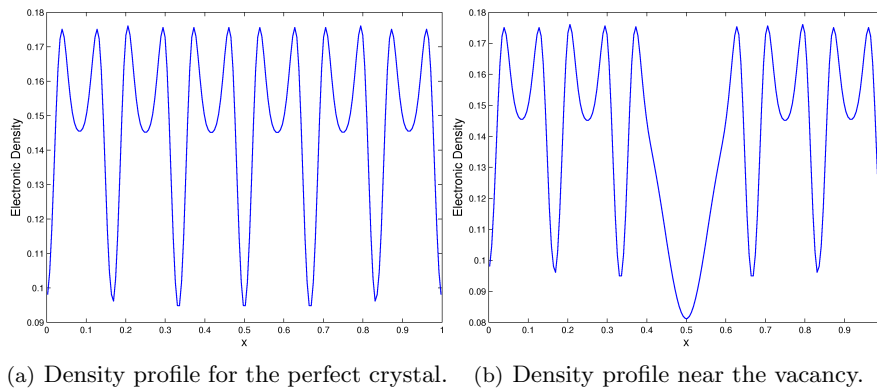


(a) Density profile for the perfect crystal.    (b) Density profile near the vacancy.

FIG. 4.4. *Electronic density across the center line in figure 4.3.*

as in the periodic case, we chose the initial density by extending periodically the density from the smooth region, and interpolating onto the grid in the non-smooth region.

The electronic density obtained with this procedure in both cases is shown in Figure 4.3. A detail of the electronic density is shown in Figure 4.4. It is seen in the figures that the matching at the interface between the smooth and non-smooth regions is seamless.

It is important to note as well that the effect on the electronic density of the disturbance in the crystalline lattice is limited to one or two periodic cells in each direction. This allows us to define a non-smooth region localized near the defect, ensuring a sub-linear scaling algorithm.

**5. The Kohn-Sham density functional theory**

Kohn-Sham DFT is quite different from TFW due to the presence of orbitals.

From a numerical viewpoint, orbitals present at least three problems: The first is that they are not uniquely defined. The second is that they in principle extend to the whole space. The third is the orthonormality constraints that are usually imposed on the orbitals.

We will work with localized orbitals. This allows us to link each orbital with particular atoms or bonds. However, localized orbitals are not uniquely defined either. This non-uniqueness issue is fixed by the particular localization procedure that we use, in the localized subspace iteration algorithm discussed below (see also [8]). Using localized orbitals also allows us to truncate the orbitals outside some localization region. This also fixes the second problem. To deal with the third problem, we use a non-orthogonal formulation of KS-DFT. In this formulation, we only need to enforce independence of the orbitals, which is guaranteed automatically by the localization procedure. Since we work with localized orbitals, we will adopt a real space formulation of the problems and sub-problems that we encounter, even for the periodic case.

The basic idea of the sub-linear scaling algorithm is as follows. We divide the orbitals into two sets: the set associated with the atoms in the smooth region and the set associated with the atoms in the non-smooth region. The orbitals in the smooth region can be approximated using asymptotic analysis, and can be found beforehand by solving a series of periodic problems. We then go back to the full problem, and use the results in the smooth region (as given) to find the orbitals in the non-smooth region. The wave functions in the smooth region serve as the environment for the wave functions in the non-smooth region. They confine the wave functions in the non-smooth region and force them to decay into the smooth region.

This decay property of the orbitals will be used in an essential way in the algorithm, both for the smooth and non-smooth regions. In the smooth region, even though the problems are periodic, we will formulate them as problems in the whole real space, in contrast to the conventional $k$-point sampling techniques for periodic problems. The real space formulation is preferred since the problem in the non-smooth region has to be formulated in real space anyway. The decay property of the wave functions then allows us to truncate the computational domain and compute the wave functions only on small neighborhoods of the corresponding physical domains. Indeed, a crucial component of our algorithm is to ensure that the wave functions we compute have fast decay properties. For this purpose, we replace the standard orthogonalization steps in KS-DFT algorithms by localization steps, as in [8].

As in the last section, we will first discuss the asymptotics for the smooth region and the general KS-DFT algorithm that we will use. We will then discuss the sublinear scaling algorithm.

**5.1. Asymptotic analysis in the smooth region.** The rescaled energy functional for the Kohn-Sham density functional theory model is

$$
I_\varepsilon(\{\psi_k\}) = \frac{\varepsilon^2}{2} \sum_k \int_{\mathbb{R}^3} |\nabla \psi_k(y)|^2 \, \mathrm{d}y + \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}(\varepsilon^3 \rho) \rho(y) \, \mathrm{d}y
$$
$$
+ \frac{\varepsilon}{2} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{(\rho - m)(y)(\rho - m)(y')}{|y - y'|} \, \mathrm{d}y \, \mathrm{d}y'. \tag{5.1}
$$

For simplicity, the nonlocal pseudopotential is not included here. Similar results can be obtained when it is included.

The associated Euler-Lagrange equations are

$$-\frac{\varepsilon^2}{2}\Delta\psi_k + V_{\mathrm{xc}}(\varepsilon^3\rho)\psi_k - \phi\psi_k + \sum_{k'}\lambda_{kk'}\psi_{k'} = 0; \tag{5.2}$$

$$-\Delta\phi = 4\pi\varepsilon(m-\rho). \tag{5.3}$$

Here the $\lambda$'s are the Lagrange multipliers for the orthonormality constraints, $\phi$ is the Coulombic potential generated by the charge distribution of electrons and ions, and $V_{\mathrm{xc}}(\varepsilon^3\rho) = \epsilon_{\mathrm{xc}}(\varepsilon^3\rho) + \epsilon'_{\mathrm{xc}}(\varepsilon^3\rho)\varepsilon^3\rho$.

The wave functions $\{\psi_k\}$ are far from being unique. We will assume that the collection of $\{\psi_k\}$ can be chosen as $\{\psi_\alpha(y_i, (x-x_i)/\varepsilon)\}$, where $\alpha$, which ranges from 1 to $n_0/2$, is the index for the occupied states. $\psi_\alpha(y,\cdot)$ is localized at 0, $i.e.$,it decays away from the origin. Thus $\psi_\alpha(y_i, (x-x_i)/\varepsilon)$ is localized at the atomic position $x_i$. If the system is in equilibrium or under homogeneous deformation, $\{\psi_\alpha(y_i, (x-x_i)/\varepsilon)\}$ can be chosen as the well-known Wannier function [41] of the $\alpha$-th energy band centered at $x_i$.

Similar to the case of the TFW model, we take the following ansatz:

$$\psi_\alpha(y,z) = \frac{1}{\varepsilon^{3/2}}\psi_{\alpha,0}\left(y,\frac{x}{\varepsilon}\right) + \frac{1}{\varepsilon^{1/2}}\psi_{\alpha,1}\left(y,\frac{x}{\varepsilon}\right) + \varepsilon^{1/2}\psi_{\alpha,2}\left(y,\frac{x}{\varepsilon}\right) + \cdots, \tag{5.4}$$

$$\rho = \rho\left(y,\frac{x}{\varepsilon}\right) = \frac{1}{\varepsilon^3}\rho_0\left(y,\frac{x}{\varepsilon}\right) + \frac{1}{\varepsilon^2}\rho_1\left(y,\frac{x}{\varepsilon}\right) + \frac{1}{\varepsilon^1}\rho_2\left(y,\frac{x}{\varepsilon}\right) + \cdots, \tag{5.5}$$

$$\phi = \phi\left(y,\frac{x}{\varepsilon}\right) = \phi_0\left(y,\frac{x}{\varepsilon}\right) + \varepsilon\phi_1\left(y,\frac{x}{\varepsilon}\right) + \varepsilon^2\phi_2\left(y,\frac{x}{\varepsilon}\right) + \cdots. \tag{5.6}$$

As discussed above, $\psi_\alpha(y,z)$ decays when $z$ becomes large, and $\rho(y,z)$ and $\phi(y,z)$ are periodic in $z$, as in the TFW model. Recall that

$$\rho(y) = 2\sum_\alpha \sum_{x_j \in \varepsilon L \cap \Omega} |\psi_\alpha\left(y_j, \frac{x-x_j}{\varepsilon}\right)|^2.$$

In the limit as $\varepsilon \to 0$, by the decay property, we obtain, at leading order, that

$$\rho_0(y,z) = 2\sum_\alpha \sum_{z_j \in L} |\psi_{\alpha,0}(y,z-z_j)|^2. \tag{5.7}$$

Note that $\psi_{\alpha,0}(y,\cdot-z_j)$ is a translation of $\psi_{\alpha,0}(y,\cdot-z_k)$ with translation vector $z_j - z_k$. Similarly, for the orthonormality constraint, we have

$$\int_{\mathbb{R}^3} \psi^*_{\alpha,0}(y,z-z_i)\psi_{\alpha',0}(y,z-z_j)\,\mathrm{d}z = \delta_{\alpha\alpha'}\delta_{ij}/\det((\boldsymbol{I}+\nabla u)(x)). \tag{5.8}$$

Taking these into consideration and expanding the Euler-Lagrange equations, we get the leading order equations:

$$-\frac{1}{2}\Delta^x_2\psi_{\alpha,0}(y,z) + V_{\mathrm{xc}}(\rho_0)\psi_{\alpha,0}(y,z) - \phi_0(y,z)\psi_{\alpha,0}(y,z)$$

$$+ \sum_{\alpha',z_j \in L}\lambda_{\alpha\alpha',z_j 0}\psi_{\alpha',0}(y,z-z_j) = 0; \tag{5.9}$$

$$-\Delta^x_2\phi_0(y,z) = 4\pi(m_0 - \rho_0)(y,z). \tag{5.10}$$

Here the $\lambda$'s are Lagrange multipliers. The second order equations are

$$
-\frac{1}{2}\Delta_2^x\psi_{\alpha,1}(y,z)-\frac{1}{2}\nabla_2^x\cdot\nabla_1\psi_{\alpha,0}(y,z)-\frac{1}{2}\nabla_1\cdot\nabla_2^x\psi_{\alpha,0}(y,z)
$$
$$
+V_{\mathrm{xc}}(\rho_0)\psi_{\alpha,1}(y,z)+V'_{\mathrm{xc}}(\rho_0)\rho_1\psi_{\alpha,0}(y,z)-\phi_0\psi_{\alpha,1}(y,z)
$$
$$
-\phi_1\psi_{\alpha,0}(y,z)+\sum_{\alpha',z_j\in L}\lambda_{\alpha\alpha',z_j0}\psi_{\alpha',1}(y,z-z_j)
$$
$$
+\sum_{\alpha',z_j\in L}\lambda_{\alpha\alpha',z_j1}\psi_{\alpha',0}(y,z-z_j)=0; \tag{5.11}
$$

$$
-\Delta_2^x\phi_{\alpha,1}(y,z)-\nabla_2^x\cdot\nabla_1\phi_{\alpha,0}(y,z)-\nabla_1\cdot\nabla_2^x\phi_{\alpha,0}(y,z) \tag{5.12}
$$
$$
=4\pi(m_1-\rho_1)(y,z).
$$

As was observed in [10], the asymptotic analysis reveals something that may seem surprising at first sight, namely that the problem becomes effective local: all terms in the asymptotic expansion are determined by solving differential equations in the fast variable; the slow variable enters only as a parameter. In the language of fiber bundles [11], all terms are determined by solving PDEs on the fibers — differentiation with respect to the variables on the base manifold only enters in the source terms. Indeed, given the displacement field, the leading order terms become independent for different values of the slow variable. This means that the local electronic structures at different macro-scale locations are essentially independent of each other and can be solved independently. This is a remarkable fact and is a consequence of the near-sightedness discovered by W. Kohn [26].

**5.2. Essential ingredients of the KS-DFT algorithm.**    To ease discussion, we will use the KS-DFT model in the following form:

$$
\min_{\{\psi_j\}}\frac{1}{2}\sum_{j=1}^{N}\int_{\mathbb{R}^3}|\nabla\psi_j|^2\,\mathrm{d}y+\sum_{j=1}^{N}\int_{\mathbb{R}^3}(V_{ps}\psi_j)(y)\psi_j(y)\,\mathrm{d}y
$$
$$
+\frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3}\frac{(\rho-m)(y)(\rho-m)(y')}{|y-y'|}\,\mathrm{d}y\,\mathrm{d}y'+\int_{\mathbb{R}^3}\epsilon_{\mathrm{xc}}(\rho)\rho\,\mathrm{d}y, \tag{5.13}
$$

subject to the orthogonality constraints

$$
\int_{\mathbb{R}^3}\psi_i\psi_j\,\mathrm{d}y=\delta_{ij}. \tag{5.14}
$$

The electronic density, $\rho$, is defined as

$$
\rho(y)=2\sum_{j=1}^{N}|\psi_j(y)|^2. \tag{5.15}
$$

Given $N_a$ atoms located at $\{\boldsymbol{R}_k\}$, $k=1,\ldots,N_a$, the ionic function, $m$, is defined by

$$
m(y)=\sum_{k=1}^{N_a}m_k^a(y-\boldsymbol{R}_k), \tag{5.16}
$$

where $m_k^a$ is the atomic function.

The last term in (5.13) is the exchange and correlation energy in the local density approximation (LDA). As in the TFW case, we use the Ceperley-Alder exchange and correlation energy, as parametrized by Perdew and Zunger [4, 35].

The pseudopotential operator in (5.13), $V_{ps}$, together with the ionic function $m$, describe the screened effect of the core electrons and the nuclei. Here we use the norm-conserving pseudopotential of Troullier and Martins [38], in the Kleinman-Bylander form [24]. The pseudopotential is usually decomposed into local and nonlocal components:

$$V_{\mathrm{ps}} = \left(V_{\mathrm{loc}} - m * \frac{1}{|y|}\right)\boldsymbol{I} + \sum_l V_{\mathrm{nonlocal},l}\widehat{P}_l, \tag{5.17}$$

where $\widehat{P}_l$ projects out the $l$-th angular momentum component of the wave function. The function $V_a$ in (5.16) is chosen so that the local part in the pseudopotential has compact support, with approximately the same cutoff radius as the nonlocal components, and so that

$$\int (\rho - m)(y)\,\mathrm{d}y = 0. \tag{5.18}$$

The minimization problem (5.13) leads to the nonlinear eigenvalue problem

$$\boldsymbol{H}\psi_j = \sum_{i,j} \epsilon_{ij}\psi_i, \tag{5.19}$$

where $\{\epsilon_{ij}\}_{i,j=1}^N$ are the Lagrange multipliers necessary to enforce the orthogonality constraint (5.14), and $\boldsymbol{H}$ is the Hamiltonian:

$$\boldsymbol{H}\psi = -\frac{1}{2}\Delta\psi + V_{\mathrm{eff}}(\rho)\psi. \tag{5.20}$$

In (5.20), $V_{\mathrm{eff}}$ is the effective potential. This problem is usually solved by diagonalization, resulting in an algorithm that scales like $O(N^3)$.

It is important to notice that the Kohn-Sham energy (5.13) is invariant under unitary transformation of the set of wave functions. In fact, one can even extend the energy functional to accommodate non-orthonormal wave functions [30].

$$\min_{\{\psi_j\}} \frac{1}{2}\sum_{j,k}\left(\int_{\mathbb{R}^3}(\nabla\psi_k)^{\mathrm{T}}(\boldsymbol{S}^{-1})_{kj}\nabla\psi_j\,\mathrm{d}y + \int_{\mathbb{R}^3}(V_{ps}\psi_k)(\boldsymbol{S}^{-1})_{kj}\psi_j\,\mathrm{d}y\right)$$

$$+ \frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3}\frac{(\rho - m)(y)(\rho - m)(y')}{|y - y'|}\,\mathrm{d}y\,\mathrm{d}y' + \int_{\mathbb{R}^3}\epsilon_{\mathrm{xc}}(\rho)\rho\,\mathrm{d}y, \tag{5.21}$$

with $\rho$ defined as

$$\rho(y) = 2\sum_{ij}\psi_i(y)(\boldsymbol{S}^{-1})_{ij}\psi_j(y). \tag{5.22}$$

Here $\boldsymbol{S}$ is the overlap matrix

$$S_{ij} = \int \psi_i\psi_j\,\mathrm{d}y.$$

In this case, the energy functional is invariant with respect to any non-singular transformation of the wave functions. This allows us to reformulate the problem. In this reformulation, the emphasis is not on a particular basis, such as the eigenfunctions considered in (5.19), but on the linear subspace generated by these functions. The starting point in our algorithm is a set of linearly independent normalized wave functions, $\Psi = \{\psi_j\}_{j=1}^N$. The wave functions are centered at fixed locations, $\{b_j\}_{j=1}^N$, which we usually take to be the center of the interatomic bond. These wave functions are compactly supported, with support within a certain cutoff radius, $R_c$. As in the work of Zhou, Saad, Tiago and Chelikowsky [46], we iteratively improve this initial subspace by polynomial filtering. But in contrast to their work, in our approach there is no orthogonalization step, which would lead to cubic scaling. Instead, linear scaling is achieved in our algorithm by using always a representation in terms of a localized basis.

Next, we discuss some algorithmic details.

As in [46], we use a Chebyshev polynomial filter. The Chebyshev polynomial of degree $n \geq 0$ is defined as

$$T_n(x) = \begin{cases} \cos(n\cos^{-1} x), & |x| \leq 1, \\ \cosh(n\cosh^{-1} x), & |x| \geq 1. \end{cases} \tag{5.23}$$

These polynomials satisfy the three-term recurrence relation

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \tag{5.24}$$

Because of the rapid growth of the polynomials outside the interval $[-1,1]$, they can efficiently filter out the components in the wave functions $\Psi$ that correspond to eigenvalues in the high end of the spectrum. Given $V = \mathrm{span}\,\Psi$, the filtered subspace is

$$\widetilde{\Psi} = p_m(\boldsymbol{H})\Psi, \tag{5.25}$$

where $p_m$ is a shifted Chebyshev polynomial, so that the unwanted part of the spectrum lies in $[-1,1]$, and the lower part of the spectrum lies in $(-\infty,-1]$. The filter is applied via the use of the recursion relation (5.24), which makes it unnecessary to construct $p_m(\boldsymbol{H})$ explicitly.

Given a set of wave functions, $\{\psi_j\}_{j=1}^N$, centered at the locations $\{b_j\}_{j=1}^N$, respectively, we obtain a more localized basis by minimizing

$$F[\varphi] = \frac{\int_{\mathbb{R}^3} |y - b_j|^{2p} |\varphi(y)|^2 \, \mathrm{d}y}{\int_{\mathbb{R}^3} |\varphi(y)|^2 \, \mathrm{d}y} \tag{5.26}$$

among functions $\varphi$ of the form

$$\varphi(y) = \sum_{k=1}^{r} \alpha_k \psi_k(y). \tag{5.27}$$

Only the wave functions whose support overlaps with the support of $\psi_j$ are included in the summation (5.27). Since the support of the wave functions is contained within a pre-defined cutoff radius from the center, there is an upper bound on the number of wave functions required in (5.27), $r$, which is independent of the number of atoms. In [8] it is proved that the localized wave functions decay faster for increasing values

of $p \in \mathbb{N}$, but the conditioning of the problem grows with $p$ as well. In what follows, we use the value $p = 3$.

Minimization of (5.26) leads to the generalized eigenvalue problem

$$\boldsymbol{W} a = \lambda \boldsymbol{S} a, \tag{5.28}$$

where $a = (\alpha_1, \ldots, \alpha_N)$, $\lambda$ is the generalized eigenvalue, and the matrices $\boldsymbol{W}$ and $\boldsymbol{S}$ are defined as

$$\boldsymbol{W}_{ij} = \int_{\mathbb{R}^3} |y - b_k|^{2p} \psi_i(y) \psi_j(y) \, \mathrm{d}y, \tag{5.29}$$

$$\boldsymbol{S}_{ij} = \int_{\mathbb{R}^3} \psi_i(y) \psi_j(y) \, \mathrm{d}y, \quad i, j = 1, \ldots, r. \tag{5.30}$$

The smallest generalized eigenvalue and corresponding eigenvector of (5.28) are obtained using the Lanczos algorithm for matrix pairs [36], and a bisection procedure to isolate the smallest eigenvalue [20].

Note that in the procedure described above the Hamiltonian $H$ is fixed, or equivalently, the electronic density is fixed. To monitor the convergence of the iterated spaces, we compute the corresponding electronic density by

$$\rho_k(y) = 2 \sum_{i,j} \psi_i^k(y) (\boldsymbol{S}^{-1})_{ij} \psi_j^k(y).$$

The LSI is repeated until there is little change in the density, at which point the Hamiltonian is updated. Therefore the algorithm consists of two iterations: an inner loop in which the linear subspace is iteratively improved via the LSI, and a self-consistency outer loop in which the electronic density, and therefore the Hamiltonian, is updated. In the algorithm described below we update the density in the self-consistent iteration using linear mixing. However, this is not a limitation of the algorithm, and other schemes may be used [29].

In summary, this linear-scaling algorithm proceeds as follows:

1: Given wave functions $\Psi_0$.
2: $k = 0$.
3: **repeat** {(*Self-Consistency Loop (SCF)*)}
4:     $l = 0$.
5:     Compute electronic density:

$$\rho_k(y) = 2 \sum_{i,j} \psi_i^k(y) (\boldsymbol{S}^{-1})_{ij} \psi_j^k(y).$$

6:     $\rho_{k,0} = \rho_k$.
7:     Compute effective potential: $V_{\mathrm{eff}}(\rho) = V_H(\rho) + V_{ps} + V_{Xc}(\rho)$.
8:     **repeat** {(*Localized Subspace Iteration (LSI) [8]*)}
9:         Filtering Step: $\widetilde{\Psi}_l = p_m(\boldsymbol{H}) \Psi_l$.
10:        Localization Step:
11:        **for** $r = 1$ to $N$ **do**
12:            Localize $\psi_r$:

$$\min_{\varphi \in \mathrm{span} \, \widetilde{\Psi}_l} \frac{\int_{\mathbb{R}^3} |y - b_j|^{2p} |\varphi(y)|^2 \, \mathrm{d}y}{\int_{\mathbb{R}^3} |\varphi(y)|^2 \, \mathrm{d}y}.$$

13:          **end for**
14:          Truncation: $\varphi_k(y) = 0$ if $\|y - b_k\| \geq R_c$, $k = 1, \ldots, N$.
15:          Update subspace: $\Psi_{l+1} = [\varphi_1, \ldots, \varphi_N]$.
16:          Compute associated density, $\rho_{k,l+1}$.
17:      **until** $\|\rho_{k,l+1} - \rho_{k,l}\|_2 \leq$ Tol.
18:      Mixing: $\rho_{k+1} = \alpha \rho_k + (1 - \alpha)\rho_{k,l+1}$.
19: **until** $\|\rho_{k+1} - \rho_k\|_2 \leq$ Tol.

The overlap matrix, $\boldsymbol{S}$, is symmetric and positive definite, and its condition number is bounded independently of its size. Due to the truncation of the wave functions, it is also sparse. Therefore the matrix can be inverted with linear complexity using the multigrid or the preconditioned conjugate gradient methods.

**5.3. Algorithm for the smooth region.**     The electronic structure in the smooth region can be computed in a similar way as for the TFW model: A coarse grid of unit cells is generated. The wave functions associated with the electrons inside these unit cells are found by solving the asymptotic equations. The multiscale interpolation technique is used to generate the electronic structure at other locations. The first and the last steps are the same as in the previous case. Let us focus on the second step. To do this, as in Section 4.4, define the periodic Kohn-Sham functional (here the nonlocal pseudo-potential is included)

$$
I_{\boldsymbol{A}}(\{\psi\}) = \frac{1}{2}\sum_\alpha \int_{\mathbb{R}^3} |(\boldsymbol{I} + \boldsymbol{A})^{-\mathrm{T}} \nabla \psi_\alpha(z; \boldsymbol{A})|^2 \, \mathrm{d}z + \sum_\alpha \int_{\mathbb{R}^3} (V_{ps}\psi_\alpha)\psi_\alpha(z; \boldsymbol{A})\, \mathrm{d}z
$$
$$
+ \int_\Gamma \epsilon_{\mathrm{xc}}(\rho)\rho(z; \boldsymbol{A})\, \mathrm{d}z + \frac{1}{2}\iint_{\Gamma \times \Gamma} (\rho - m_{\mathrm{CB}})(z; \boldsymbol{A})G(z - z'; \boldsymbol{A})(\rho - m_{\mathrm{CB}})(z'; \boldsymbol{A})\, \mathrm{d}z\, \mathrm{d}z',
$$
$$(5.31)$$

where $m_{\mathrm{CB}}(z; \boldsymbol{A}(x)) = m_0(y, z) = \sum_{z_i} m_0^a((\boldsymbol{I} + \boldsymbol{A})(x)(z - z_i))$, and $G(z; \boldsymbol{A})$ is the periodic Coulombic kernel as in (4.23). It is easy to see that the Euler-Langrange equations of $I_{\boldsymbol{A}}(\{\psi\})$ with the orthonormality constraints

$$
\int_{\mathbb{R}^3} \psi_{\alpha,0}^*(z - z_i; \boldsymbol{A})\psi_{\alpha,0}(z - z_j; \boldsymbol{A})\, \mathrm{d}z = \delta_{\alpha\alpha'}\delta_{ij}/\det(\boldsymbol{I} + \boldsymbol{A}) \tag{5.32}
$$

are (5.9)-(5.10) with $\boldsymbol{A} = \nabla u(x)$. Therefore, we can obtain the leading order approximation of the wave functions in the smooth region using the LSI algorithm described above.

Although this looks like an orthogonal formulation of DFT, the numerical algorithm that we use is the LSI algorithm described in last subsection, and that is a non-orthogonal formulation. The orthogonal formulation used here and in Sec. 5.1 is only for the convenience of presenting the asymptotic analysis.

When solving for the wave functions in one unit cell, we have to consider their periodic translations in the other unit cells. Since the wave functions are localized, we only have to consider a small number of these translations. We choose the cutoff radius to be $R_c = 1.5a_0$, where $a_0$ is the lattice constant.

In this formulation, the wave functions are not periodic. However, the electronic density is periodic. Therefore the Coulomb term is computed as in the TFW model, *i.e.,* by solving the Poisson equation

$$
-\Delta\phi = 4\pi(\rho - m), \tag{5.33}
$$

with periodic boundary conditions, discussed in Sec. 4.4.

Some remarks are in order concerning the strategy used here to solve the unit cell problems in the smooth region. Notice that the cell problem is not formulated as a periodic problem over the unit cell, but rather a problem over the whole space, as shown in (5.31). The wave function, although localized, is also defined on all of $\mathbb{R}^3$.

We further remark that, although (5.31) is derived from assumptions of periodicity, it takes a form that is quite different from the usual formulation of periodic models when wave functions are included (*e.g.*, for the Hartree and Hartree-Fock models as in [2, 3]). Our formulation (5.31) is in real space, not in $k$-space where the Bloch boundary condition is used [29]. Because of this, we have chosen a localized representation (Wannier functions) for the electronic states, since they are more naturally associated with each individual atom or electron. For numerical computation, the real space formulation might be more advantageous than the $k$-space representation. This point will be pursued in future work.

**5.4. Algorithm for the non-smooth region.**    Let $K$, $K_s$ and $K_{ns}$ denote the collection of indices for the wave functions associated with the whole domain, atoms in the smooth region and atoms in the non-smooth region, respectively. Note that $K = K_s \cup K_{ns}$, and $K_s \cap K_{ns} = \emptyset$. As before, the electronic structure in the non-smooth region is solved by minimizing the original energy functional, but with the orbitals associated with the atoms in the smooth region already given:

$$\inf_{\{\psi_k\}_{k \in K_{ns}}} \frac{1}{2} \sum_{k \in K} \sum_{j \in K} \left( \int_{\mathbb{R}^3} (\nabla \psi_k)^{\mathrm{T}} (\boldsymbol{S}^{-1})_{kj} \nabla \psi_j \, \mathrm{d}y + \int_{\mathbb{R}^3} (V_{ps}\psi_k)(\boldsymbol{S}^{-1})_{kj} \psi_j \, \mathrm{d}y \right)$$
$$+ \int_{\mathbb{R}^3} \epsilon_{\mathrm{xc}}(\rho)\rho(y) \, \mathrm{d}y + \frac{1}{2} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{(\rho - m)(y)(\rho - m)(y')}{|y - y'|} \, \mathrm{d}y \, \mathrm{d}y'. \quad (5.34)$$

Here, we have adopted the non-orthogonal formulation, so that $\rho$ is defined as

$$\rho(y) = 2 \sum_{ij} \psi_i(y)(\boldsymbol{S}^{-1})_{ij} \psi_j(y), \quad (5.35)$$

and satisfies the normalization condition:

$$\int_{\mathbb{R}^3} \rho(y) dy = N. \quad (5.36)$$

The problem is to determine the electronic structure for the non-smooth region given the environment of the smooth region, *i.e.*, given $\{\psi_k\}_{k \in K_s}$, we need to find $\{\psi_k\}_{k \in K_{ns}}$.

At this point, the problem is formulated over the entire space. The orbitals are defined everywhere. However, due to the localization of the orbitals, we can truncate the integrals in the definition of the functional (5.34) as well as the inverse overlap matrix.

We define an environment region, with indices $K_{\mathrm{env}} \subset K_s$. This environment region contains the indices of the wave functions that overlap with the non-smooth region after the wave functions are truncated, as shown in Figure 5.1. The density can therefore be written as

$$\rho(y) = 2 \sum_{i \in K_{ns} \cup K_{\mathrm{env}}} \sum_{j \in K_{ns} \cup K_{\mathrm{env}}} \psi_i (\mathbf{S}^{-1})_{ij} \psi_j(y)$$
$$+ 4 \sum_{i \in K_{ns} \cup K_{\mathrm{env}}} \sum_{j \in K_s \setminus K_{\mathrm{env}}} \psi_i (\mathbf{S}^{-1})_{ij} \psi_j(y) + 2 \sum_{i \in K_s \setminus K_{\mathrm{env}}} \sum_{j \in K_s \setminus K_{\mathrm{env}}} \psi_i (\mathbf{S}^{-1})_{ij} \psi_j(y).$$
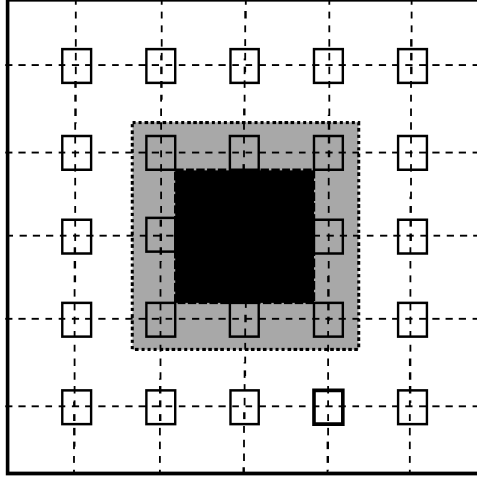$$(5.37)$$

Fig. 5.1. *Decomposition of the domain into smooth and non-smooth regions. The non-smooth region is shown in black, and the environment region is shown in dark gray.*

This decomposition is exact. Furthermore, if $y \in \Omega_{ns}$, only the first term in (5.37) is not zero:

$$\rho(y) = 2 \sum_{i,j \in K_{ns} \cup K_{env}} \psi_i (\mathbf{S}^{-1})_{ij} \psi_j(y), \quad y \in \Omega_{ns}. \tag{5.38}$$

At first sight, since $\boldsymbol{S}$ is a huge matrix of the dimension of the total number of wave functions in the system, it is very expensive to invert $\boldsymbol{S}$ or to calculate $\boldsymbol{S}^{-1} v$ using the conjugate gradient method. The localization of the wave functions helps again, since the overlap matrix can be truncated away from a small neighborhood of the non-smooth region. Therefore, we define the density in the non-smooth region as

$$\rho_{ns}(y) = 2 \sum_{i,j \in K_{ns} \cup K_{env}} \psi_i (\widetilde{\mathbf{S}}^{-1})_{ij} \psi_j(y), \quad y \in \Omega_{ns}, \tag{5.39}$$

where we have truncated the overlap matrix and only consider the wave functions in the non-smooth region and the environment region:

$$\widetilde{\mathbf{S}}_{ij} = \int_{\mathbb{R}^3} \psi_i(y) \psi_j(y) \, dy, \quad i, j \in K_{ns} \cup K_{env}. \tag{5.40}$$

This is the main approximation we make that helps to decouple the orbitals in the non-smooth region from a majority of the orbitals in the (vast) smooth region.

Let $y(\Omega'_{ns})$ be a domain that contains the support of the orbitals associated with atoms in the non-smooth region. The functional in (5.34) can now be simplified to:

$$\inf_{\{\psi_k\}_{k \in K_{ns}}} \frac{1}{2} \sum_{k \in K} \sum_{j \in K} \left( \int_{\mathbb{R}^3} (\nabla \psi_k)^{\mathrm{T}} (\boldsymbol{S}^{-1})_{kj} \nabla \psi_j \, \mathrm{d}y + \int_{\mathbb{R}^3} (V_{ps} \psi_k)(\boldsymbol{S}^{-1})_{kj} \psi_j \, \mathrm{d}y \right)$$
$$+ \int_{y(\Omega'_{ns})} \epsilon_{\mathrm{xc}}(\rho) \rho(y) \, \mathrm{d}y + \frac{1}{2} \iint_{y(\Omega'_{ns}) \times \mathbb{R}^3} \frac{(\rho - m)(y)(\rho - m)(y')}{|y - y'|} \, \mathrm{d}y \, \mathrm{d}y'. \tag{5.41}$$
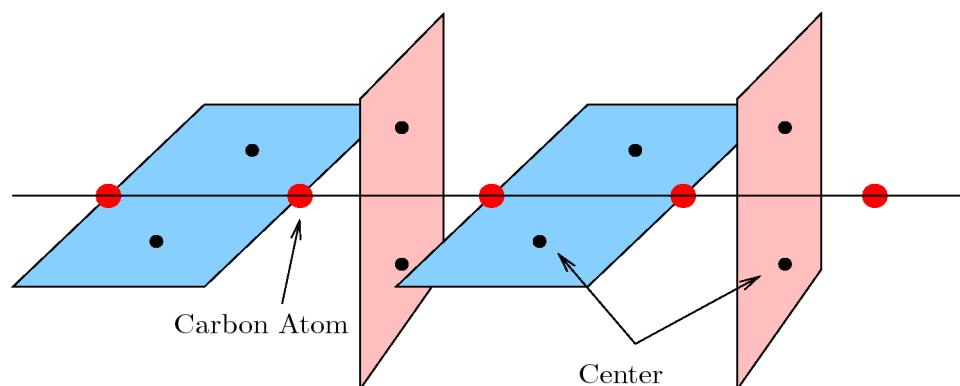
FIG. 5.2. *Carbon Chain. In red, the location of the atoms. The wave functions have centers on alternating orthogonal planes.*

There are several interaction terms that have to be considered. The first is the Coulomb term. This situation is identical to the TFW case, and we treat the long range interactions (like the nonlocal pseudopotential) in the same way. Other than these long range interactions, the overlap between the wave functions in the smooth region and the wave functions in the non-smooth region also needs to be taken into consideration. It may seem at first sight that to evaluate (5.41), for terms like

$$\int_{\mathbb{R}^3} (\nabla \psi_k)^{\mathrm{T}} (\boldsymbol{S}^{-1})_{kj} \nabla \psi_j \, \mathrm{d}y,$$

all the wave functions should be included, which will make the calculation impossible. Fortunately, since we are working with localized wave functions, only a small number of wave functions in the smooth region need to be taken into account.

The LSI algorithm discussed above is not an optimization algorithm (such as steepest decent or quasi-Newton methods). Instead an effective Hamiltonian is constructed using the current electron density

$$\boldsymbol{H} = -\frac{1}{2}\Delta + V_{\mathrm{eff}}(\rho_{ns}) + V_H(\rho_s), \tag{5.42}$$

which is used to get an updated set of localized wave functions. In this algorithm, the surrounding wave functions from the smooth region contribute in two ways: On one hand, they are involved when the overlap matrix is calculated in the definition of the density. On the other hand, the surrounding wave functions serve to confine the wave functions in the non-smooth region and force them to decay away from the non-smooth region through the localization procedure.

This coupling between the smooth and non-smooth regions via the wave functions is not standard, but it is the most natural approach one can think of for this situation. In this approach, the non-smooth region sees the surrounding wave functions as their environment, and this reflects the situation for real materials.

Finally, here we have used LSI to solve the Kohn-Sham model. This is not the only choice. Other algorithms based on localized wave functions such as orbital minimization can also be used.

**5.5. Numerical examples.**     To illustrate the linear scaling algorithm presented in Sec. 5.2, we have considered an infinite carbon chain. Each unit cell contains two atoms. There are two covalent bonds between every two consecutive atoms in the chain, and therefore we associate four wave functions to each unit cell. Note that these wave functions are not confined to their unit cell; only the centers, which are used in the localization step, are. The atoms are located on the $OX$ axis, and the wave functions are centered at points located on alternating orthogonal planes, as depicted in Figure 5.2. The distance between atoms is $a_0 = 1.785$Å. We use 4 unit cells in the $OX$ direction. The environment region contains 3 unit cells in the $OX$ axis on each side of the non-smooth region. The truncation radius for the wave functions is $R_c = 1.5a_0$. We set the tolerance for both the inner and outer loop to $10^{-6}$. The density was updated every 100 iterations within the inner loop. We have used $64 \times 32 \times 32$ gridpoints in each unit cell. The electronic density obtained by the algorithm presented in Sec. 5.2 is shown in Figure 5.3.

## 6. Conclusions

We have presented a new class of sub-linear scaling algorithms for studying the electronic structure of materials. To better appreciate the features of our approach, let
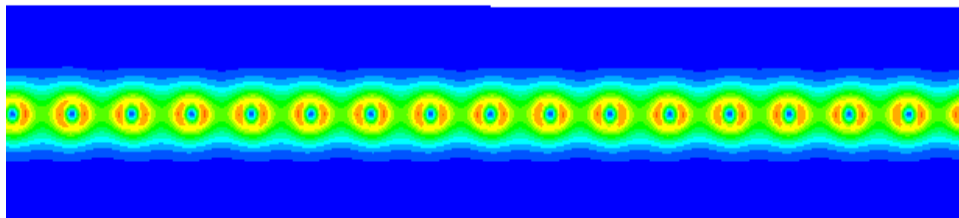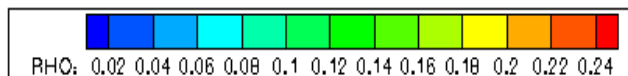


(a) carbon chain — no vacancy. In this case, we still artificially divide the domain into a smooth and non-smooth region, in order to compare with the results for the case with vacancy (see below).
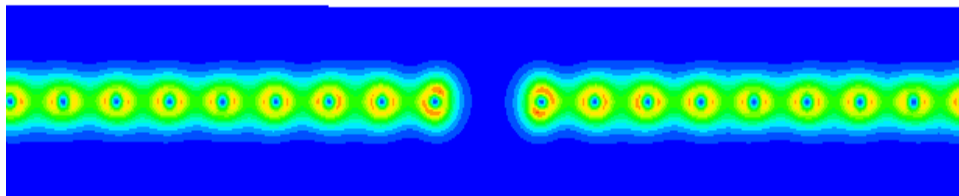


(b) carbon chain with a vacancy. The non-smooth region covers a neighborhood of the vacancy. Both smooth and non-smooth regions are shown.

FIG. 5.3. *Electronic density of a carbon chain, computed with the linear scaling algorithm presented in Sec. 5.2.*

us recapitulate the main steps for the electronic structure analysis using Kohn-Sham density functional theory. The main components are:

- Partition the set of orbitals into a union of two sets: the orbtials associated with the atoms in the region where the deformation of the material is smooth, and the orbitals associated with atoms in the region around defects.

- The (localized) orbitals associated with atoms in the smooth region are computed appoximately by solving the asymptotic equations.

- The localized orbitals in the non-smooth region are found by minimizing the total energy subject to the condition that the wave functions associated with the electrons in the smooth region (the "environment wave functions") are given. Due to the fast decay properties of the wave functions, only a small number of environment wave functions need to be taken into account.

There are two main approximations in this procedure: The first is the asymptotic solutions in the smooth region. The second is the truncation of the set of "environment wave functions" away from the non-smooth region. Both approximations can be systematically improved to obtain better and better accuracy. In particular, we have avoided introducing an ad hoc buffer region.

The general idea pursued here, that of solving simplified asymptotic equations in the smooth region, has an obvious similarity with the quasicontinuum method [25, 37]. For one thing, the leading order equation is nothing but the Cauchy-Born approximation. There are some differences though, compared with the quasicontinuum method. The numerical methods that we use for the simplified equations are not limited to the finite element method. In particular, the finite difference method offers some clear advantage: it fits naturally to the common crystal structures such as the body-centered cubic, face-centered cubic and diamond structures. Its flexibility can be explored to reduce the error at the boundary of the non-smooth region, the analog of the "ghost force". We should note that such an error does not arise for the electronic structure problems considered here. But it will be a concern when we study the deformation of the material.

Aside from this main theme, some of the ideas we introduced along the way might also be of interest. For example, we introduced a real space formulation and algorithm for the electronic structure of a unit cell. In the past this problem has been usually formulated and solved in $k$-space. We also introduced a formulation for the electronic structure of a subsystem, giving that the electronic structure of the rest of the system is known. We feel this is the natural formulation for the electronic structure problem with a Dirichlet boundary condition, and could potentially be useful for other purposes.

In this paper we have limited our attention to the electronic structure problem. The natural next step is to study the deformation of the material itself, as was done in [17] for the TFW model. This will be done in future work.

## REFERENCES

[1] X. Blanc, C. Le Bris and P.L. Lions, *From molecular models to continuum mechanics*, Arch. Ration. Mech. Anal., 164(4), 341–381, 2002.

[2] I. Catto, C. Le Bris and P.L. Lions, *On the thermodynamic limit for Hartree-Fock type models*, Ann. Inst. Henri Poincaré, 18(6), 687–760, 2001.

[3] I. Catto, C. Le Bris and P.L. Lions, *On some periodic Hartree-type models for crystals*, Ann. Inst. Henri Poincaré, 19(2), 143–190, 2002.

[4] D.M. Ceperley and B.J. Alder, *Ground state of the electron gas by a stochastic method*, Phys. Rev. Lett., 45(7), 566–569, 1980.

[5] N. Choly, G. Lu, W. E and E. Kaxiras, *Multiscale simulations in simple metals: a density-functional- based methodology*, Phys. Rev. B, 71(9), 94–101, 2005.

[6] W. E, B. Engquist, X. Li, W. Ren and E. Vanden-Eijnden, *Heterogeneous multiscale methods: a review*, Commun. Comput. Phys., 2(3), 367–450, 2007.

[7] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1(1), 87–132, 2003.

[8] W. E, T. Li and J. Lu, *Localized basis of eigen-subspaces and operator compression*, submitted, 2007.

[9] W. E, J. Lu and J. Yang, *Uniform accuracy of the quasicontinuum method*, Phys. Rev. B, 74(21), 214115, 2006.

[10] W. E and J. Lu, *The continuum limit and quasi-continuum approximation of quantum mechanics models for crystalline solids*, Commun. Math. Sci., 5(3), 679–696, 2007.

[11] W. E and J. Lu, *Seamless multiscale modeling via dynamics on fiber bundles*, Commun. Math. Sci., 5(3), 649–663, 2007.

[12] M. Fago, R.L. Hayes, E.A. Carter and M. Ortiz, *Density functional theory based local quasicontinuum method: Prediction of dislocation nucleation*, Phys. Rev. B, 70, 100–102, 2004.

[13] G. Galli, *Linear scaling methods for electronic structure calculations and quantum molecular dynamics simulations*, Curr. Opin. Solid State Mater. Sci., 1(6), 864–874, 1996.

[14] J. Gao, *Methods and applications of combined quantum mechanical and molecular mechanical potentials*, Reviews in Computational Chemistry, VCH Publishers, New York, 7, 119–185, 1995.

[15] C.J. García-Cervera, J. Huang, J. Lu and W. E, *Numerical algorithms for treating the Coulomb term in density functional theory*, in preparation, 2007.

[16] C.J. García-Cervera, *An efficient real space method for orbital-free density-functional theory*, Commun. Comput. Phys., 2(2), 334–357, 2007.

[17] V. Gavini, K. Bhattacharya and M. Ortiz, *Quasi-continuum orbital-free density-functional theory: a route to multi-million atom non-periodic DFT calculation*, J. Mech. Phys. Solids, 55(4), 697–718, 2007.

[18] V. Gavini, J. Knap, K. Bhattacharya and M. Ortiz, *Non-periodic finite-element formulation of orbital-free density functional theory*, J. Mech. Phys. Solids, 55(4), 669–696, 2007.

[19] S. Goedecker, *Linear scaling electronic structure methods*, Rev. Mod. Phys., 71(4), 1085–1123, 1999.

[20] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Third edition, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996.

[21] D. Gottlieb and S.A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications*, CBMS-NSF Regional Conference Series in Appl. Math., 26, SIAM, Philadelphia, 1977.

[22] L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comp. Phys., 73(2), 325–348, 1987.

[23] E. Isaacson and H.B. Keller, *Analysis of Numerical Methods*, John Wiley & Sons, Inc., New York, 1966.

[24] L. Kleinman and D.M. Bylander, *Efficacious form for model pseudopotentials*, Phys. Rev. Lett., 48(20), 1425–1428, 1982.

[25] J. Knap and M. Ortiz, *An analysis of the quasicontinuum method*, J. Mech. Phys. Solids, 49(9), 1899–1923, 2001.

[26] W. Kohn, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett., 76(17), 3168–3171, 1996.

[27] E.H. Lieb and B. Simon, *The Thomas-Fermi theory of atoms, molecules and solids*, Adv. Math., 23(1), 22–116, 1977.

[28] G. Lu, E.B. Tadmor and E. Kaxiras, *From electrons to finite elements: A concurrent multi-scale approach for metals*, Phys. Rev. B, 73(2), 024108, 2006.

[29] R. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge Univ. Press, Cambridge, 2004.

[30] F. Mauri, G. Galli and R. Car, *Orbital formulation for electronic-structure calculations with linear system-size scaling*, Phys. Rev. B, 47(15), 9973–9976, 1993.

[31] D. Negrut, M. Anitescu, A. El-Azab and P. Zapol, *Quasicontinuum-like reduction of DFT calculations of nanostructures*, J. Nanosci. Nanotechnol., 2007.

[32] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer-Verlag, New York, 1999.

[33] H.J. Nussbaumer, *Fast Fourier Transform and Convolution Algorithms*, Springer Series in Information Sciences, Springer-Verlag, Berlin, 2, 1981.

[34] R.G. Parr and W.Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford Univ. Press, New York, 1989.

[35] J.P. Perdew and A. Zunger, *Self interaction correction to density functional approximations for many electron systems*, Phys. Rev. B, 23(10), 5048–5079, 1981.

[36] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Algorithms and Architectures for Advanced Scientific Computing, Manchester University Press, Manchester, 1992.

[37] E.B. Tadmor, M. Ortiz and R. Phillips, *Quasicontinuum analysis of defects in solids*, Philos. Mag. A, 73, 1529–1963, 1996.

[38] N. Troullier and J.L. Martins, *Efficient pseudopotentials for plane-wave calculations*, Phys. Rev. B, 43(3), 1993–2006, 1991.

[39] L.W. Wang and M.P. Teter, *Kinetic-energy functional of the electron density*, Phys. Rev. B, 45(23), 13196–13220, 1992.

[40] Y.A. Wang, N. Govind and E.A. Carter, *Orbital-free kinetic-energy density functionals with a density-dependent kernel*, Phys. Rev. B, 60(24), 16350–16358, 1999.

[41] G.H. Wannier, *The structure of electronic excitation levels in insulating crystals*, Phys. Rev., 52(3), 191–197, 1937.

[42] A. Warshel,*Computer Modeling of Chemical Reactions in Enzymes and Solutions*, Wiley-Interscience, New York, 1991.

[43] W. Yang, *Direct calculation of electron density in density-functional theory*, Phys. Rev. Lett., 66(11), 1438–1441, 1991.

[44] X. Zhang, G. Lu, *An orbital-free density function theory based QM/MM methodology for metals*, preprint.

[45] Y. Zhang, H. Liu and W. Yang, *Ab initio QM/MM and free energy calculations of enzyme reactions*, Computational Methods for Macromolecules - Challenges and Applications, T. Schlick, H. Gan, Lect. Notes Comput. Sci. Eng., Springer, New York, 332–354, 2002.

[46] Y. Zhou, Y. Saad, M.L. Tiago and J.R. Chelikowsky, *Self-consistent-field calculations using Chebyshev-filtered subspace iteration*, J. Comp. Phys., 219(1), 172–184, 2006.