Sparsity of the Density Matrix in Kohn-Sham Density Functional Theory and an Assessment of Linear System-Size Scaling Methods

Roi Baer and Martin Head-Gordon

Department of Chemistry, University of California, Berkeley, California 94720, and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

(Received 22 April 1997)

The range and sparsity of the one-electron density matrix (DM) in density functional theory is studied for large systems using the analytical properties of its Chebyshev expansion. General estimates of the range of the DM are derived, showing that the range is inversely proportional to the square root of an insulator band gap and inversely proportional to the square root of the temperature. These findings support "principle of nearsightedness" introduced recently by W. Kohn [Phys. Rev. Lett. **76**, 3168 (1996)]. These estimates are used to study the complexity of several linear system-size scaling electronic structure algorithms which differ in their dependence on the geometric dimensionality of the system. [S0031-9007(97)04543-2]

PACS numbers: 71.15.Mb

The Hohenberg-Kohn and Kohn-Sham (KS) density functional theory (DFT) of the many-electron ground state energy has served as a basis for numerous large-scale electronic-structure and molecular dynamics simulations, especially using plane-wave total energy and Car-Parrinello approaches [1,2]. These methods are, however, capable of dealing with a limited number of atoms, presently on the order of 10^2, primarily due to the $O(MN_e^2)$ scaling, where N_e is the number of electrons in a unit cell and M is the number of plane waves. A great challenge thus remains for dealing with systems containing many more atoms which have no simplifying symmetry such as lattice periodicity. Recently, it was pointed out [3–5] (and see also Ref. [6] for earlier ideas) that by invoking a nonorthogonal basis of localized functions, it is possible to develop methods which scale linearly with system size. Pursuing this idea, several algorithms for electronic-structure calculations have been developed [7–20].

Recently, Kohn [20] introduced the principle of "the nearsightedness of equilibrium systems," with which he argues that unlike the KS orbitals, the one-electron density matrix (DM) $\rho(r, r')$ is always short ranged in $|r - r'|$, thereby enabling a linear system-size scaling.

This paper attempts to contribute to these efforts by introducing a new analytical method to derive estimates for the range of the DM in an electronic system, based on a few characteristics of the system. This allows us to compare the intrinsic complexity of several algorithmic approaches and evaluate their scaling with the accuracy of the calculation and the dimensionality of the system. The method also demonstrates in a simple way Kohn's nearsightedness principle concerning the range of the DM. The technique makes use of the Chebyshev polynomial expansion of the DM introduced recently by Goedecker and Colombo [14,15].

KS-DFT reduces the $2N_e$ electron problem to a one-electron self-consistent field (SCF) procedure leading to global energy minimization. The SCF procedure involves two steps combined and iterated to convergence. The first is the construction of a Hamiltonian $\hat{H} = -\hbar^2/2m_e\nabla^2 + \hat{V}_{\text{eff}}$, where m_e is the electron mass and the effective potential $V_{\text{eff}}(\mathbf{r})$ is built from the electron density $\rho(\mathbf{r})$. The second step is the calculation of the electron density determined by the Hamiltonian, given in the Kohn-Sham scheme as $\rho(\mathbf{r}) = \sum_i n_i|\psi_i(\mathbf{r})|^2$, where $\psi_i(\mathbf{r})$ are normalized eigenfunctions of the Hamiltonian $\hat{H}\psi_i = \varepsilon_i\psi_i$ and n_i is an occupancy number, equal to 1 if $\varepsilon_i \le \varepsilon_{N_e}$ and 0 otherwise.

New theoretical and numerical developments [7,8] have recently enabled the step of constructing the Hamiltonian from the density to scale linearly with system size. This can be achieved for both KS-DFT and Hartree-Fock effective Hamiltonians [9]. With these advances, the computational bottleneck shifts to the second stage, where simple diagonalization schemes lead to expensive $O(N^3)$ scaling.

It is convenient at this point to define the DM by

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i n_i\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}'). \tag{1}$$

Note that the DM is a projector onto the space spanned by the N_e lowest energy orbitals, so an alternative formulation based on the Heaviside weight $\theta(\varepsilon)$, equal to 1 if ε is positive and to 0 if ε is negative, is

$$\hat{\rho} = \theta(\mu - \hat{H}). \tag{2}$$

The chemical potential μ is determined by the electron number: $\text{tr}\,\hat{\rho} = N_e$. For systems with band gap $\delta\varepsilon$, the DM can be approximated to precision 10^{-D} by the Fermi-Dirac (FD) matrix (FDM):

$$\hat{F} = \frac{1}{1 + e^{\beta(\hat{H}-\mu)}}, \tag{3}$$

where the chemical potential μ is determined by the condition $\text{tr}\,\hat{F} = N_e$ and the inverse temperature parameter β

 0031-9007/97/79(20)/3962(4)$10.00

controls the accuracy, chosen so that discrepancies larger than 10^{-D} between the FD and the Heaviside functions are confined within the band gap and do not affect the corresponding DMs. The relation of β to the gap $\delta\varepsilon$ should be

$$\beta\delta\varepsilon/2 \geq D\log 10\,. \tag{4}$$

This equivalence of the FDM and the insulator DM (to the precision D) is due to the existence of a band gap. For conductors, Eq. (3) can still be used at low temperatures [21].

Unlike $\theta(\varepsilon)$, the function $F(\varepsilon) = (1 + e^{\beta(\varepsilon-\mu)})^{-1}$ is infinitely differentiable on the real line. This allows efficient use of polynomial expansions to approximate the DM in numerical applications as first suggested by Goedecker and Colombo [14]:

$$F(\hat{H}) = \sum_{n=0}^{P-1} a_n(\beta,\mu)p_n(\hat{H})\,. \tag{5}$$

Here, p_n are orthogonal polynomials (with respect to some weight) and a_n are numerical constants, dependent on the temperature and chemical potential.

The choice of the polynomial series for the expansion is crucial for good convergence. A Taylor expansion around any energy value does not converge at all due to poles of the FD function in the complex plane limiting the radius of convergence. Following Goedecker *et al.* [14,15], we choose Chebyshev polynomials which converge fastest in the maximum norm [22]. Chebyshev expansions have proved invaluable in quantum dynamical computations since their introduction to that field by Kosloff and Tal-Ezer [23] (see Refs. [24–27] for examples and developments) and recently have also been used to directly expand the DM in plane-wave DFT calculations [28].

We now briefly summarize the Chebyshev expansion of the FDM. The expansion is written as

$$F(\hat{H}) = \sum_{n=0}^{P-1} a_n(\beta_s,\mu_s)T_n(\hat{H}_s)\,, \tag{6}$$

where P is the expansion length and $\hat{H}_s$ is a shifted and scaled Hamiltonian, constructed so its eigenvalues are contained in the interval $[-1,1]$. To be specific, we define $E_{\max}$ and $E_{\min}$ as the largest and smallest eigenvalues of $\hat{H}$, and

$$\hat{H}_s = \frac{\hat{H} - \overline{E}}{\Delta E}\,, \tag{7}$$

where

$$\overline{E} = \frac{E_{\max} + E_{\min}}{2}; \qquad \Delta E = \frac{E_{\max} - E_{\min}}{2}\,. \tag{8}$$

Similarly, the scaled inverse temperature and a scaled-shifted chemical potential are

$$\beta_s = \beta\Delta E, \qquad \mu_s = (\mu - \overline{E})/\Delta E\,. \tag{9}$$

$T_n(x) = \cos(n\cos^{-1}x)$ is the nth Chebyshev polynomial, and the expansion coefficients are

$$a_n(\beta_s,\mu_s) = \frac{2}{(1+\delta_{n0})\pi} \times \int_{-1}^{1} \frac{T_n(x)}{\sqrt{1-x^2}}\,\frac{1}{1+e^{\beta_s(x-\mu_s)}}\,dx\,. \tag{10}$$

The error committed in using a finite expansion is smooth and uniform throughout the interpolated interval. In a future publication, we show that when $2\pi \ll \sqrt{1-\mu_s^2}\,\beta_s$ (a condition invariably met in practice), the required length P of the expansion for an accuracy 10^{-D} $(D > 1)$ is largely insensitive to μ_s and closely estimated by

$$P \approx \tfrac{2}{3}(D-1)\beta_s\,. \tag{11}$$

This equation generalizes a relation introduced by Goedecker and Colombo [14].

Combining Eq. (4) with Eq. (11), the length of the Chebyshev expansion for insulating systems is determined by

$$P = 3D(D-1)\frac{\Delta E}{\delta\varepsilon} \tag{12}$$

(here, the numerical constants multiply to a value of ≈ 3).

We exploit the fact that the Chebyshev expansion is of finite length for studying properties of the DM, and we now estimate its spatial range $W(\hat{F})$. The Hilbert space associated with the system can be represented using a finite basis of normalized Gaussians $G_{\mathbf{r}}$ of range σ, centered on a mesh of points $\mathbf{r}$, with spacing a comparable in size to σ. The mesh points are in a large box containing the system. The overlap matrix is

$$S_{\mathbf{r}',\mathbf{r}} = \langle G_{\mathbf{r}'} \mid G_{\mathbf{r}}\rangle = e^{-(\mathbf{r}-\mathbf{r}')^2/2\sigma^2}, \tag{13}$$

and the dual biorthonormal basis can be defined by

$$\langle\overline{G}_{\mathbf{r}}| = \sum_{\mathbf{r}'}(S^{-1})_{\mathbf{r}\mathbf{r}'}\langle G_{\mathbf{r}'}|\,. \tag{14}$$

We state results for a finite basis and then take the limit to an infinite basis by indefinitely decreasing the mesh spacing a and the width of the Gaussians σ (keeping a/σ constant). We assume that for a large separation, a basis function and its dual have similar functional behavior, and the Hamiltonian matrix elements take the form

$$\langle\overline{G}_{\mathbf{r}'}|\hat{H}|G_{\mathbf{r}}\rangle \approx e^{-(\mathbf{r}-\mathbf{r}')^2/2\sigma^2}. \tag{15}$$

The prefactor of the exponent has been dropped, since due to the locality of the interactions, it depends only weakly on $|\mathbf{r}-\mathbf{r}'|$ when the latter is large. The long-range matrix elements for $\hat{H}^2$ can be estimated by a Gaussian composition rule as

$$\langle\overline{G}_{\mathbf{r}'}|\hat{H}^2|G_{\mathbf{r}}\rangle = \sum_{\mathbf{x}}\langle\overline{G}_{\mathbf{r}'}|\hat{H}|G_{\mathbf{x}}\rangle\langle\overline{G}_{\mathbf{x}}|\hat{H}|G_{\mathbf{r}}\rangle \approx e^{-(\mathbf{r}-\mathbf{r}')^2/4\sigma^2}. \tag{16}$$

Thus, the range of $\hat{H}^2$ is $\sqrt{2}\,\sigma$, and repeatedly using the Gaussian composition rule, the range of H^P is estimated as $\sqrt{P}\,\sigma$.

Now, since the Chebyshev expansion of the FDM involves operating with the Hamiltonian P times, we conclude that the spatial range of the DM is also of the order $\sqrt{P}\,\sigma$. Special cancellations due to the coefficients of the expansion may *reduce* the range further, so this estimate is probably an upper bound in most cases. Using Eq. (12) for the expansion length P, the range of the DM for insulators is then given by

$$W(\hat{F}) = \sqrt{\frac{3\sigma^2\Delta E}{\delta\varepsilon}\,D(D-1)}\,. \tag{17}$$

This equation contains two representation-dependent parameters: the spatial width σ of the basis functions and the eigenvalue range of the Hamiltonian matrix, given by $\Delta E = (E_{\max} - E_{\min})/2$. For small enough σ, $E_{\min}$ is, at worst, dominated by the minimal values of the potential energy on mesh points close to atomic centers, where due to the Coulomb potential $E_{\min} \propto \sigma^{-1}$. (Actually, $E_{\min}$ is finite even when $\sigma \to 0$.) $E_{\max}$ is the kinetic energy of a Gaussian; thus $E_{\max} \approx \hbar^2\sigma^{-2}/2m_e$. As σ diminishes, the kinetic energy term dominates ΔE, and Eq. (17) converges to

$$W(\hat{F}) \approx \sqrt{\frac{3\hbar^2}{4m_e\delta\varepsilon}\,D(D-1)}\,. \tag{18}$$

This result conforms to Kohn's estimates for the range of orbitals in one dimensional periodic systems [29]. The arguments we present can be considered a generalization of Kohn's theorem to systems of any dimension. Furthermore, for nonperiodic systems, although Eq. (18) probably overestimates the range of the DM, it establishes a finite range for it, a conclusion derived also in Refs. [30,31].

The range of the DM for a finite temperature is computed using Eq. (11), with physical temperature $(k_B\beta)^{-1}$:

$$W(\hat{F}) \approx \sqrt{P}\,\sigma = \sqrt{\frac{\hbar^2}{3m_e}\,(D-1)\beta}\,. \tag{19}$$

This estimate is independent of the chemical potential so it parallels a free electron gas model for metals. For insulators, it overestimates the FDM range unless the temperature is very high. Equation (19) closely resembles the thermal de Broglie wavelength $\lambda = \sqrt{h^2/3m_e k_B T}$, which serves to estimate the range of DM correlations for a particle in thermal equilibrium. The dependence on the square root of $D-1$ shows that initially the density matrix decays faster than an exponential of the distance, as is actually the case for the range of the DM of a homogeneous gas of noninteracting electrons.

We now discuss the DM sparsity of a large system in a finite basis, composed of N functions with a Hamiltonian matrix $H_{mn} = \langle m|\hat{H}|n\rangle$. We assume that the system is large so the Hamiltonian is very sparse. In order to measure sparseness, we introduce the concept of *breadth*. For a given accuracy 10^{-D}, the breadth of a column vector $\mathbf{v}$, $B_D(\mathbf{v})$ is the number of its elements having a magnitude larger than 10^{-D}. For the breadth of a matrix H, $B_D(H)$ is the maximal breadth of its columns. Based on the preceding results, we conclude that $B_D(\hat{H}^n) \approx n^{d/2}B_D(\hat{H})$, where d is the geometric dimensionality of the system. The DM breadth is then

$$B_D(\hat{F}) \approx P^{d/2}B_D(\hat{H})\,, \tag{20}$$

where P is given by Eq. (12) for ground state calculations or by Eq. (11) for finite temperature calculations, and $B_D(\hat{H})$ is the Hamiltonian breadth, primarily determined by the breadth of the inverse overlap matrix S^{-1}. The latter breadth is proportional to the *square root* of the condition number of the overlap matrix [32].

Using the breadth given in Eq. (20), it is possible to assess the algorithmic complexity of several linear scaling approaches. We first discuss methods that calculate the DM by minimization of a modified energy functional constructed to ensure idempotency. We name these "$F \times F$ methods." The minimization process consists of a sequence of computations of a power of the DM F^n, where $n = 2, 3, 4$ in the Li-Nunes-Vanderbilt-Daw (LNV-D) [11–13], Hernandez *et al.* [17,33], and Kohn [20] approaches, respectively. The total numerical work in the $F \times F$ methods is then

$$J[F \times F] \approx M_n P^d B_D(\hat{H})^2 N\,, \tag{21}$$

where M_n is the number of $F \times F$ evaluations required for minimization until F is determined to 10^{-D} in some norm. Next, we discuss the computation of the DM by a Chebyshev expansion [15]. The computation now consists of applying $\hat{H}_s$ P times to N unit vectors, so

$$J[\text{Cheb}] \approx P^{d/2+1}B_D(\hat{H})^2 N\,. \tag{22}$$

In order to demonstrate that these relations are indeed reflected in actual calculations, both Chebyshev and LNV-D methods were applied to a tight-binding cubic lattice model, having 10^d sites ($d = 1, 2, 3$ is the dimensionality) and a nearest neighbor spacing of 4 a.u. The parametrization of the model was based on the Hamiltonian of Ref. [34] for carbon, but the following changes were made for simplifying the interactions, achieving a large band gap and smaller spectral range: Only two electrons were allocated to each atom and separately for each dimension d we changed the magnitude of the Slater-Koster parameter $V_{ss,\sigma}$ until a band gap of $\delta\varepsilon = 0.1$ a.u. was achieved and the spectral range ΔE was in the range 0.25–0.3 a.u. We plotted in Fig. 1 the numerical work J as a function of the theoretical polynomial length P determined by the accuracy D [via Eq. (11)], where

$$10^{-D} = \max\{|\mathrm{tr}(\rho^2 - \rho)|, \sqrt{\mathrm{tr}([\rho, H_s]^2)}\}/\mathrm{tr}(\rho)\,. \tag{23}$$

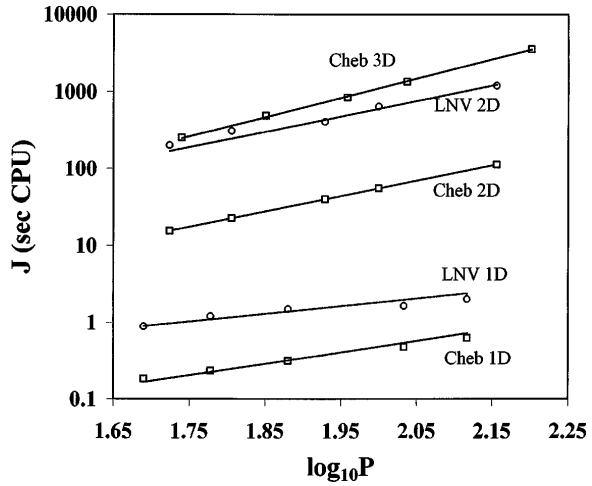

FIG. 1. The numerical work vs Chebyshev polynomial length. Circles (LNV-D) and squares (Chebyshev) are calculation results, while lines are of slope given by equations in the text. LNV-D calculations for a 3D system were not attempted due to memory limitations.

Both methods were applied using a sparse matrix tree code [32]. In the LNV-D implementation, the conjugate gradients method was used for minimizing the energy, starting from $F = 1/2\hat{\mathbf{I}}$. Also plotted in the figure are lines with the slope taken from Eqs. (21) and (22). It can be seen that, in general, the trends shown by the actual numerical work, with respect to the dependence on dimensionality and accuracy, correspond satisfactorily to the theory presented above. Figure 1 and Eqs. (21) and (22) show that different linear scaling methods can have different complexities depending on the geometry of the system.

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley Laboratory under U.S. DOE Contract No. DE-AC03-76SF00098. R. B. thanks D. Neuhauser for helpful discussions. M. H. G. acknowledges support from Packard (1995-2000).

[1] M. C. Payne, M. P. Teter, D. C. Allen, T. A. Arias, and J. D. Joanopolus, Rev. Mod. Phys. **64**, 1045 (1992).
[2] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[3] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[4] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).
[5] W. Kohn, Chem. Phys. Lett. **208**, 167 (1993).
[6] P. W. Anderson, Phys. Rev. Lett. **21**, 13 (1968).
[7] C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon, Chem. Phys. Lett. **230**, 8 (1994); *ibid.* **253**, 268 (1996).
[8] M. C. Strain, G. E. Scuseria, and M. J. Frisch, Science **271**, 51 (1996).
[9] E. Schwegler, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[10] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[11] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10 891 (1993).
[12] M. S. Daw, Phys. Rev. B **47**, 10 895 (1993).
[13] R. W Nunes and D. Vanderbilt, Phys. Rev. B **50**, 17 611 (1994).
[14] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994).
[15] S. Goedecker and M. Teter, Phys. Rev. B **51**, 9455 (1995).
[16] E. B. Stechel, A. R. Williams, and P. J. Feibelman, Phys. Rev. B **49**, 10 088 (1994).
[17] E. Hernadez and M. J. Gillan, Phys. Rev. B **51**, 10 157 (1995).
[18] P. Ordejon, D. A. Drabold, R. M. Martin, and M. P. Grumbach, Phys. Rev. B **51**, 1456 (1995).
[19] P. Ordejon, E. Artacho, and J. Soler, Phys. Rev. B **53**, 10 441 (1996).
[20] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[21] A. K. Rajagopal, Adv. Chem. Phys. **41**, 59 (1980).
[22] T. J. Rivlin, *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory* (Wiley, New York, 1990), 2nd ed.
[23] R. Kosloff and H. Tal-Ezer, Chem. Phys. Lett. **127**, 223 (1986).
[24] D. Neuhauser and M. Baer, J. Chem. Phys. **90**, 4351 (1989).
[25] R. Baer and R. Kosloff, Chem. Phys. Lett. **200**, 183 (1992).
[26] D. K. Hoffman, Y. Huang, W. Zhu, and D. J. Kouri, J. Chem. Phys. **101**, 1242 (1994).
[27] R. Baer, Y. Zeiri, and R. Kosloff, Phys. Rev. B **54**, 5287 (1996).
[28] D. J. Kouri, Y. Huang, and D. K. Hoffman, J. Phys. Chem. **100**, 7903 (1996).
[29] W. Kohn, Phys. Rev. **115**, 809 (1959).
[30] W. Kohn and R. J. Onffroy, Phys. Rev. B **8**, 2485 (1973).
[31] A. Nenciu and G. Nenciu, Phys. Rev. B **47**, 10 112 (1993).
[32] R. Baer and M. Head-Gordon (to be published).
[33] E. Hernadez, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **53**, 7147 (1996).
[34] A. P. Horsfield, P. D. Godwin, D. G. Pettifor, and A. P. Sutton, Phys. Rev. B **54**, 15 773 (1996).