

Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms

W. Kohn

Department of Physics, University of California, Santa Barbara, California 93106

(Received 21 August 1995)

A widely applicable “nearsightedness” principle is first discussed as the physical basis for the existence of computational methods scaling linearly with the number of atoms. This principle applies to the one particle density matrix $n(r, r')$ but not to individual eigenfunctions. A variational principle for $n(r, r')$ is derived in which, by the use of a penalty functional $P[n(r, r')]$, the (difficult) idempotency of $n(r, r')$ need not be assured in advance but is automatically achieved. The method applies to both insulators and metals.

PACS numbers: 71.15.Mb

An aspect which gives current versions of density function theory (DFT) great practical importance is that for systems of N atoms, with $N \gg 1$, the computational effort scales as N^2 to N^3 , while traditional configuration interactions methods in practice scale with much higher powers of N ($\sim N^7$) and, indeed, asymptotically exponentially as $\sim e^{\alpha N}$. The latter methods are currently limited to $N \leq 5-10$, while DFT methods have recently been able to deal with $N \sim 100-200$.

In the last several years there has been a great deal of interest in developing so-called $O(N)$ methods for DFT, i.e., methods which scale linearly with N [1-6]. They promise to allow calculations, within the next few years, of systems consisting of 10^3 to 10^5 atoms.

The present paper makes a contribution to this effort. Unlike some other recent work, it does not depend on the existence of well-localized generalized Wannier functions, which exist only in large-gap insulators. It applies to both insulators and metals.

I first discuss a widely applicable physical principle which explains why $O(N)$ methods can exist. I call this principle the *nearsightedness* of equilibrium systems consisting of *many* quantum mechanical particles moving in an external potential $v(r)$. For systems without long range electric fields it can loosely be expressed as follows: Let $F(r_1, r_2, \dots, r_\nu)$ be a static property depending on ν coordinates r_1, \dots, r_ν , all within a restricted volume ω of linear dimension $\sim \lambda$, a typical de Broglie wavelength occurring in the ground state wave function or finite T ensemble. [The density $n(r)$ and pair-correlation function $g(r, r')$ are examples.] Denote by \bar{r} the center of mass, $\bar{r} \equiv \nu^{-1} \sum_1^\nu r_\mu$. Then, at fixed chemical potential μ , a change of the external potential $\Delta v(r')$, no matter how large, has a small effect on F , provided only that $\Delta v(r')$ is limited to a *distant* region, in the sense that for all r' , $|r' - \bar{r}| \gg \lambda$. Thus F does not “see” $\Delta v(r')$ if r' is far.

A few remarks about this principle: (1) The principle is generally a consequence of wave-mechanical destructive interference. It requires the presence of *many* particles, which need not be interacting. (2) It is not universally valid. Among exceptions are systems of noninteracting 3D bosons below the critical point, when the lowest one-

particle wave function has a macroscopic occupancy, and systems with translationally invariant long range order, like a Wigner crystal in a torus. (3) The principle is tacitly assumed in much of chemistry and materials science. (4) When there are long range electric fields, as in ionic crystals, they must be self-consistently added to the external potential but do not otherwise affect nearsightedness. (A more complete account of nearsightedness will be published elsewhere.)

I will now show how nearsightedness implies the possibility of $O(N)$ methods. Take a system enclosed in a large volume Ω . Cover this volume by *overlapping* cubes of volume $\omega' = (m\lambda)^3$ where m is, say 100, in such a way that every volume ω ($\sim \lambda^3$) lies inside of, and far from the boundary of, at least one of the covering cubes. Clearly the required number of cubes is $\propto (\Omega/\omega') \propto N$. Because of nearsightedness, I can now “cut away” all but one ω' and enclose the latter in a hard wall boundary, to calculate $F(r_1, \dots, r_\nu)$. Since the number of covering cubes is $O(N)$, this may be regarded as an existence theorem for the possibility of calculating all $F(r_1, \dots, r_\nu)$, for “compact” sets of coordinates, with a computing effort linear in N .

Widely used cluster methods and the divide-and-conquer scheme [1] are quite close in spirit to the above considerations. But there are other $O(N)$ methods which do not divide the physical space Ω occupied by the system into smaller compartments, ω . These methods can be categorized as either generalized Wannier function (GWF) methods [2-5] or one-particle density matrix (DM) methods [6]. This paper is in the second category.

In the Kohn-Sham (KS) self-consistent equations [7] there occurs the following sequence:

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + v(r) \right\} \varphi_j(r) = \epsilon_j \varphi_j(r), \quad j = 1, \dots, N, \quad (1)$$

$$n(r) = \sum_1^N |\varphi_j(r)|^2, \quad (2)$$

$$E = \sum_1^N \epsilon_j. \quad (3)$$

Here N is the number of electrons, $v(r)$ is the effective one-body potential, φ_j and ϵ_j are the KS orthonormal single particle wave functions, ϵ_j are the KS eigenvalues, $n(r)$ is the particle density, and E is the total *single particle* energy. The solution of (1) is the most time-consuming step behaving as $O(N^2) - O(N^3)$. We can characterize the sequence (1)–(3) by the scheme

$$v(r) \rightarrow \{\varphi_j(r), \epsilon_j\} \rightarrow \left\{ n(r), E \left(\equiv \sum_1^N \epsilon_j \right) \right\}. \quad (4)$$

Now it is well known that to calculate $n(r)$ and E we do not need the *individual* φ_j , ϵ_j but only the one-particle density matrix

$$n(r, r') \equiv \sum_{j=1}^N \varphi_j^*(r) \varphi_j(r'). \quad (5)$$

We see by inspection that

$$n(r) = n(r, r) \quad (6)$$

and

$$E = E[n] \equiv \frac{1}{2m} \int [\nabla_r \nabla_{r'} n(r, r')|_{r'=r} + v(r)n(r)] dr. \quad (7)$$

We shall develop a variational method which allows the *direct* calculation of $n(r, r')$ without the time-consuming calculation of the individual φ_j and ϵ_j ; i.e., we shall replace the scheme (4) by

$$v(r) \rightarrow \{n(r, r')\} \rightarrow n(r), E. \quad (4')$$

The DM $n(r, r')$ is short ranged in $|r - r'|$ and also nearsighted in the sense that, for fixed chemical potential μ , it depends only on the values of the potential “near” r and r' . This fact is responsible for the $O(N)$ character of the method.

From Eq. (5) we note at once the properties of $n(r, r')$

$$H \text{ Hermitian : } n(r', r) = n(r, r'), \quad (8)$$

$$N \text{ normalized : } \int n(r, r) dr = \int n(r) dr = N, \quad (9)$$

$$I \text{ idempotent : } n^2(r, r') = n(r, r'). \quad (10)$$

Here, and in what follows, expressions like $n^2(r, r')$ are matrix products,

$$n^2(r, r') \equiv \int n(r, \bar{r}) n(\bar{r}, r') d\bar{r}. \quad (11)$$

In what follows we find it more appropriate and convenient to replace the DM of Eq. (5) corresponding to fixed N by

$$n(r, r') = \sum_{\epsilon_j \leq \mu} \varphi_j^*(r) \varphi_j(r'), \quad (12)$$

where μ is the chemical potential. Thus, we replace

condition N , Eq. (9), above by

$$C \text{ chemical potential : } \epsilon_j \leq \mu. \quad (13)$$

The connection between N and μ is $\epsilon_N \leq \mu < \epsilon_{N+1}$. If N is specified rather than μ , the calculation must be repeated for several values of μ to determine which value yields N .

Any trial matrix $\tilde{n}(r, r')$ satisfying condition H above can be diagonalized and written as

$$\tilde{n}(r, r') = \sum_{j=1}^{\infty} \tilde{\lambda}_j \tilde{\varphi}_j^*(r) \tilde{\varphi}_j(r'), \quad (14)$$

where the $\tilde{\lambda}_j$ are the real eigenvalues of \tilde{n} . To satisfy also the idempotency condition, I , Eq. (10), requires

$$\tilde{\lambda}_j^2 = \tilde{\lambda}_j, \quad \text{i.e., } \tilde{\lambda}_j = 0 \text{ or } 1, \quad (15)$$

for all $\tilde{\lambda}_j$. In principle (15) can be assured by diagonalizing \tilde{n} and then replacing all λ_j by either 1 or 0. But this is unacceptably time consuming [$O(N^2) - O(N^3)$]. Below we shall derive a variational method, which will automatically lead to the satisfaction of conditions I and C above, without explicitly imposing them on the trial functions $\tilde{n}(r, r')$.

We restrict ourselves throughout to Hermitian trial functions, $\tilde{n}(r, r')$. We then form the functional

$$Q_{\mu\alpha}[\tilde{n}(r, r')] \equiv E[\tilde{n}] - \mu N[\tilde{n}] + \alpha P[\tilde{n}], \quad (16)$$

where $\tilde{n} \equiv \tilde{n}^2$, a non-negative matrix with eigenvalues $\tilde{\lambda}^2 \geq 0$; the functional E is defined in Eq. (7), where

$$N[\tilde{n}] \equiv \int \tilde{n}(r) dr, \quad (17)$$

$$P[\tilde{n}] \equiv \left[\int \tilde{n}^2(1 - \tilde{n})^2 \Big|_{r'=r} dr \right]^{1/2}; \quad (18)$$

μ is a real parameter (which plays the role of the chemical potential) and α is a positive number which will be discussed below. In terms of the eigenvalues $\tilde{\lambda}_j$ of \tilde{n} ,

$$P[\tilde{n}] = \left[\sum_1^{\infty} \tilde{\lambda}_j^2 (1 - \tilde{\lambda}_j)^2 \right]^{1/2}, \quad (18')$$

which shows that P vanishes only if every term in (18') vanishes, i.e., only if all $\tilde{\lambda}_j$ are either 0 or 1; otherwise, it is a positive *penalty functional* for violating the idempotency condition.

For orientation consider temporarily only idempotent \tilde{n} , with $P[\tilde{n}] = 0$. Then

$$Q_{\mu,\alpha}[\tilde{n}] = Q_{\mu}[\tilde{n}] = E[\tilde{n}] - \mu N[\tilde{n}] = \sum \tilde{\lambda}_j^2 (\tilde{\epsilon}_j - \mu), \quad (19)$$

where all $\tilde{\lambda}_j$ are either 0 or 1. Clearly, for a given set $\tilde{\epsilon}_j$, the minimum is attained when $\tilde{\lambda}_j = 1$ or 0 for $\tilde{\epsilon}_j \leq$ or $> \mu$, respectively. Now a well-known generalization of the Rayleigh-Ritz variational principle states that if $\tilde{\varphi}_j$

($j = 1, \dots, N$) are N orthogonal functions, with $\tilde{\epsilon}_j$ being the expectation values of the Hamiltonian, and φ_j , ϵ_j are the exact eigenfunctions and eigenvalues, then

$$\sum_1^N \tilde{\epsilon}_j \geq \sum_1^N \epsilon_j. \quad (20)$$

From this it follows straightforwardly that

$$\min_{\tilde{\varphi}_j, \tilde{\lambda}_j = 0 \text{ or } 1} \sum_1^{\infty} \tilde{\lambda}_j (\tilde{\epsilon}_j - \mu) = \sum_1^{N(\mu)} (\epsilon_j - \mu), \quad (21)$$

where $N(\mu)$ was defined following Eq. (13). Thus the first two terms in (16) favor occupying states with energies $\tilde{\epsilon}_j$ or ϵ_j below μ .

We return to the full "penalized" energy functional $Q_{\mu, \alpha}[\tilde{n}]$ of Eq. (16). We shall establish that, if α exceeds a critical α_c ,

$$Q_{\mu, \alpha}[\tilde{n}] \geq E[\tilde{n}] - \mu N[\tilde{n}] = \sum_{\epsilon_j \leq \mu} (\epsilon_j - \mu); \quad (22)$$

that equality holds only for the correct $\tilde{n} = n$; and that idempotency is automatically achieved in the minimization process.

We shall analyze Q in terms of the eigenvalues and eigenfunctions of \tilde{n} ,

$$Q_{\mu, \alpha}[\tilde{n}] = \tilde{\mathcal{E}}_{\mu}(\{\tilde{\varphi}_j, \tilde{\lambda}_j\}) + \alpha P(\{\tilde{\lambda}_j\}), \quad (23)$$

where

$$\mathcal{E}_{\mu}(\{\tilde{\varphi}_j, \tilde{\lambda}_j\}) \equiv \sum \tilde{\lambda}_j^2 (\tilde{\epsilon}_j - \mu), \quad (24)$$

$$P(\{\tilde{\lambda}_j\}) = \left[\sum_1^{\infty} \tilde{\lambda}_j^2 (1 - \tilde{\lambda}_j)^2 \right]^{1/2}. \quad (25)$$

Now consider the minimization process in two stages. Define first the conditional minimum, for which $P[\tilde{n}] = P'$,

$$\begin{aligned} \mathcal{E}_{\mu}(P') &= \min_{P[\tilde{n}] = P'} \mathcal{E}_{\mu}[\tilde{n}] \\ &= \min_{\tilde{\lambda}_j, \tilde{\varphi}_j; P(\{\tilde{\lambda}_j\}) = P'} \left[\sum \tilde{\lambda}_j^2 (\tilde{\epsilon}_j - \mu) \right]. \end{aligned} \quad (26)$$

We shall first show several properties of the minimizing $\tilde{\varphi}_j$ and $\tilde{\lambda}_j$, for given P' (≥ 0).

1. $\tilde{\lambda}_j$ is a nonincreasing function of $\tilde{\epsilon}_j$. (If false, one could interchange a pair of offending $\tilde{\lambda}_j$, lower \mathcal{E} without changing P' .)

2. The minimizing $\tilde{\varphi}_j$ are the correct eigenfunctions φ_j . (This follows from the generalization of the Rayleigh-Ritz principle that for a nonincreasing set $\{\tilde{\lambda}_j\}$, $\sum \tilde{\lambda}_j \tilde{\epsilon}_j \geq \sum \tilde{\lambda}_j \epsilon_j$ [8].)

3. For all j with $\tilde{\epsilon}_j < \mu$, $\tilde{\lambda}_j \geq 1$. [If false, replace an offending $\tilde{\lambda}_j (< 1)$ by the unique value $\xi > 1$, satisfying $\xi^2(1 - \xi)^2 = \tilde{\lambda}_j^2(1 - \tilde{\lambda}_j)^2$, which would lower \mathcal{E} without changing P' .]

4. For all j with $\tilde{\epsilon}_j > \mu$, $\tilde{\lambda}_j = 0$. (If false, replace an offending $\tilde{\lambda}_j$ by zero and increase $\tilde{\lambda}_1$ to the unique

value $\zeta > \tilde{\lambda}_1$, such that P' is not changed, while \mathcal{E} is decreased.)

5. If, for some j , $\tilde{\epsilon}_j - \mu = 0$, then $\tilde{\lambda}_j$ can be chosen arbitrarily as either 0 or 1. [If $\tilde{\lambda}_j$ were unequal to 0 or 1, replace it by 0 or 1 and increase $\tilde{\lambda}_1$ to $\lambda'_1 (> \tilde{\lambda}_1)$ reducing P' , while \mathcal{E} is unchanged.]

6. $E(P')$ is a decreasing function of P' . (Let $\{\tilde{\lambda}'_j\}$ correspond to P' . If $P'' > P'$, increase $\tilde{\lambda}'_1$ to yield P'' . The resulting \mathcal{E} is lower.)

7. Calculation of $\{\tilde{\lambda}_j\}$ for given $\{\tilde{\epsilon}_j\}$, μ , and P' . This requires minimization of the energy $\sum \tilde{\lambda}_j^2 (\tilde{\epsilon}_j - \mu)$ with respect to $\tilde{\lambda}_j$, subject to the condition $\sum \tilde{\lambda}_j^2 (1 - \tilde{\lambda}_j)^2 = P'^2$. Traditional use of a Lagrange parameter yields, on using 3–5 above,

$$\tilde{\lambda}_j = \begin{cases} \frac{1}{4}\{3 + [1 + \gamma(\mu - \tilde{\epsilon}_j)]^{1/2}\}, & \epsilon_j < \mu, \\ 0 \text{ or } 1 & \epsilon_j = \mu, \\ 0 & \epsilon_j > \mu, \end{cases} \quad (27)$$

where $\gamma (> 0)$ must be and can be uniquely chosen to yield P' .

8. General behavior of $\mathcal{E}(P')$. It follows from 7 by elementary calculations that $\mathcal{E}(P')$ is a differentiable function of P' with the following behavior:

$$\text{All } P' : \quad d\mathcal{E}(P')/dP' < 0, \quad (28)$$

$$P' \rightarrow 0 : \quad \mathcal{E}(P') = \mathcal{E}(0)$$

$$- 2P' \left[\sum_{\epsilon_j \leq \mu} (\epsilon_j - \mu)^2 \right]^{1/2}, \quad (29)$$

$$P' \rightarrow \infty : \quad \mathcal{E}(P') \rightarrow -P' \left[\sum_{\epsilon_j \leq \mu} (\epsilon_j - \mu)^2 \right]^{1/2}. \quad (30)$$

Having established properties of $E(P')$, we can now easily derive the variational principle (22) for the functional

$$Q_{\mu, \alpha}[\tilde{n}] = \min_{P'} \left\{ \min_{P[\tilde{n}] = P'} Q_{\mu, \alpha}[\tilde{n}] \right\} = \min_{P'} [\mathcal{E}(P') + \alpha P']. \quad (31)$$

Now let

$$\alpha_c = \max_{P'} \left| \frac{d\mathcal{E}(P')}{dP'} \right|;$$

then for any $\alpha > \alpha_c$, $d[\mathcal{E}(P') + \alpha P']/dP' > 0$ for all P' so that

$$\min_{\tilde{n}} Q_{\mu, \alpha}[\tilde{n}] = \mathcal{E}(0) = \sum_{\epsilon_j \leq \mu} (\epsilon_j - \mu) \quad (32)$$

(see Fig. 1). Thus, for given μ and $\alpha > \alpha_c$, minimization of $Q_{\mu, \alpha}[\tilde{n}]$, *without the imposing of the idempotency constraint in advance*, leads automatically to the correct $\tilde{n}(r, r')$, density, and energy.

Of course α_c must be at least equal to $|(d\mathcal{E}/dP')_{P'=0}| = 2[\sum_{\epsilon_j \leq \mu} (\epsilon_j - \mu)^2]^{1/2}$ which is of the order of the spectral width of the occupied levels times $N(\mu)^{1/2}$ (see Fig. 1). We surmise that, at least

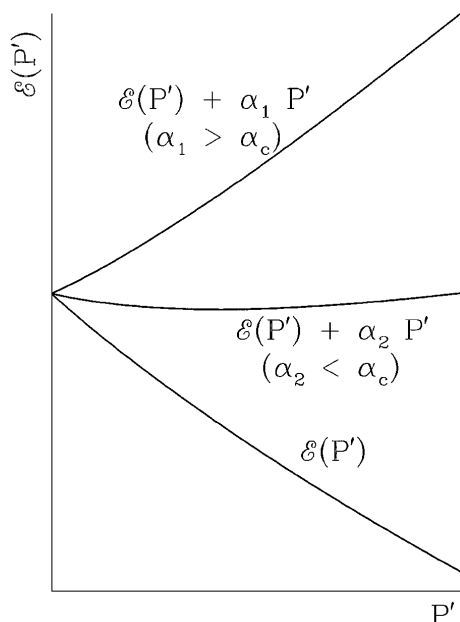


FIG. 1. The minimum energy, $\mathcal{E}(P')$, for given idempotency violation P' ; the idempotency penalty function $\alpha P'$, with $\alpha > \alpha_c$; and the penalized energy, $\mathcal{E}(P') + \alpha P'$, whose minimum occurs for $P' = 0$ and is the correct energy $E(\mu)$.

in most cases, α_c will be equal to this value but cannot yet rule out that it could be larger. Since the ϵ_j are not known in advance, one must guess α_c . If the guess was too low, the minimum of Q with respect to \tilde{n} will occur

for a positive P' . If this happens α must be increased. However, too large an α , by overemphasizing $P[\tilde{n}]$, will slow convergence.

We are currently doing an exploratory calculation in a Gaussian basis, using conventional conjugate gradient minimization with respect to the expansion coefficients of $\tilde{n}(r, r')$. For many-atom systems of unknown structure the Car-Parrinello method [9] applied to these coefficients appears to be appropriate. The high powers of \tilde{n} needed are troublesome but, we hope, not prohibitive.

I thank Professor R. Car, Professor R. Martin, Professor D. Vanderbilt, and Professor W. T. Yang for helpful discussions, and the Institute for Theoretical Physics, ETH, Zürich, for hospitality. This work was supported by Grant No. NSF-DMR93-08011.

-
- [1] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991); W. Yang and T.-S. Lee, J. Chem. Phys. **103**, 5674 (1995).
 - [2] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).
 - [3] W. Kohn, Chem. Phys. Lett. **208**, 167 (1993).
 - [4] P. Ordejon *et al.*, Phys. Rev. B **51**, 1456 (1995).
 - [5] F. Mauri *et al.*, Phys. Rev. B **47**, 9973 (1993).
 - [6] X. P. Li *et al.*, Phys. Rev. B **47**, 10891 (1993).
 - [7] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).
 - [8] E. K. U. Gross *et al.*, Phys. Rev. A **37**, 2805 (1988).
 - [9] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).