

Stationary nature of the density-functional free energy: Application to accelerated multiple-scattering calculations

D. M. C. Nicholson, G. M. Stocks, Y. Wang, and W. A. Shelton
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6114

Z. Szotek and W. M. Temmerman
Daresbury Laboratory, Daresbury, Warrington WA4 4AD, United Kingdom
(Received 30 March 1994; revised manuscript received 22 July 1994)

The number of operations required for conventional density-functional algorithms grows as the cube of the number of atoms, N . For large systems the computing requirements are unattainable. To overcome this limitation it is acceptable to approximate those variables with respect to which the free energy is stationary. We show that the stationarity of the free energy with respect to electron density, one-electron potential, chemical potential, occupation function, and temperature allows for very useful approximations leading to rapid and accurate determination of the free energy. Here we discuss approximations involved in calculating the finite temperature electron density needed to evaluate the Harris-Foulkes free energy. Of particular importance are (1) an electron density at each site that is based on exact solution of the Poisson equation combined with a solution of the multiple-scattering problem in which only scattering from a small cluster of sites surrounding the site in question is retained and (2) an approximate occupation function having a finite number of poles in the complex energy plane. The intention is to develop, within density-functional theory, an $O(N)$ scalable first-principles scheme, based on spatially local multiple-scattering methods, for calculating free energies of large systems.

A central problem in applying density-functional-theory (DFT) methods involves the $O(N^3)$ divergence in the amount of work associated with the linear-algebra operations of inverting or diagonalizing a matrix whose size is proportional to N . The ultimate objective of pursuing the following methodology is to prepare the ground for the development of scalable first-principles DFT algorithms for computing the free energy F of large systems with a number of floating point operations that grows linearly with the number of atoms, N , in the basic simulation region.¹

Within DFT, the determination of F is conveniently separated into two steps. First, the variables upon which F depends are determined, namely, electron density $\rho(\mathbf{r})$, one-electron potential $v(\mathbf{r})$, and chemical potential μ . Second, the free-energy functional must be evaluated for the particular values of $\rho(\mathbf{r})$, $v(\mathbf{r})$, and μ determined in the first step. Here we argue that the first step can be performed using real-space methods. We do not address the use of real-space methods to perform the second step. We concentrate on preparing the foundation for such a procedure and discuss two devices that allow all calculations to be performed away from the real axis in the region of the complex energy plane where real-space methods converge. This is achieved by working at finite temperature where the poles in the occupation function are separated from the real axis. Furthermore, working at fixed chemical potential eliminates the need for calculations near the real axis for the purpose of updating the chemical potential.

We focus on two approximations of interest within the context of multiple-scattering methods. First, the electron density is approximated by a sum of locally self-consistent site densities $\rho_M(\mathbf{r}_i)$, whose determination is ignorant of any scattering beyond a cluster of M atoms surrounding the i th site. Second, the occupation function $f(\epsilon - \mu)$ is approximated by $f_P(\epsilon - \mu)$, which we refer to as a band-Fermi func-

tion, where the label P is the number of its poles in the upper half of the complex energy plane. We demonstrate that, by calculating the free energy at an artificially high temperature, one is able to speed up calculations without losing the ability to accurately determine the ground-state energy.

In order to demonstrate that real-space methods are capable of providing values of $\rho(\mathbf{r})$, $v(\mathbf{r})$, and μ that are of sufficient quality to yield a free energy of the required accuracy, we perform all calculations for fcc Cu, for which the exact values are known. Furthermore, for this system the second step of evaluating the free energy can be carried out exactly for the approximate $\rho(\mathbf{r})$, $v(\mathbf{r})$, and μ . Specifically, the Korringa-Kohn-Rostoker (KKR) method is used to evaluate the eigenvalue sum and electron-hole (eh) entropy contributions to F , both of which require accurate determination of the density of states. We can therefore monitor the effect of using approximate values for $\rho(\mathbf{r})$, $v(\mathbf{r})$, and μ . To reiterate, although local densities of states at the center of each cluster of size M could have been calculated along with $\rho_M(\mathbf{r}_i)$ using real-space methods, we use exact answers obtained by conventional methods, in our case the KKR method. These results can then act as a base line. In this way, we can critically monitor the errors resulting exclusively from approximations to $\rho(\mathbf{r})$, $v(\mathbf{r})$, and μ . A detailed discussion of the convergence properties of the eigenvalue sum and eh entropy and the possibility of obtaining a completely scalable algorithm for F based on real-space methods is postponed to a further communication.

In what follows, we will utilize several stationary properties of F . These are generally valid, independent of the method used to solve the Kohn-Sham equations; however, we have in mind the multiple-scattering approach, which gives directly the electronic Green function.

In DFT the free energy is given by^{2,3}

$$F[\rho] = \int_{-\infty}^{\infty} d\epsilon \epsilon f(\epsilon - \mu) n[\epsilon, v[\rho, \mathbf{r}]] - \int d\mathbf{r} \rho v[\rho, \mathbf{r}] + U[\rho] + E_{xc}[\rho] + \mu \left(\sum_i Z_i - \int_{-\infty}^{\infty} d\epsilon f(\epsilon - \mu) n[\epsilon, v[\rho, \mathbf{r}]] \right) + k_B T \int_{-\infty}^{\infty} d\epsilon n[\epsilon, v[\rho, \mathbf{r}]] \{ f(\epsilon - \mu) \ln f(\epsilon - \mu) + [1 - f(\epsilon - \mu)] \ln [1 - f(\epsilon - \mu)] \}, \quad (1)$$

where $n[\epsilon, v[\rho, \mathbf{r}]]$ is the density of states for $v[\rho, \mathbf{r}]$, $f(\epsilon - \mu)$ is the occupation function, and Z_i is the nuclear charge at the site i . The terms on the right-hand side of Eq. (1) are usually referred to as the eigenvalue sum (first term), the ‘‘double counting’’ contribution (next three) [which includes the Coulomb (U) and exchange-correlation (E_{xc}) energies], and eh entropy (last), respectively. The remaining term measures the energy cost of failing to choose the chemical potential in such a way that the number of occupied electron levels is equal to the number of protons. In Eq. (1), $v[\rho, \mathbf{r}]$ is the so-called input potential; it is related to the electron density through the Schrödinger equation in the following manner:

$$(-\nabla^2 + v[\rho, \mathbf{r}] - \epsilon_i) \Psi_i = 0 \quad (2)$$

and

$$\rho(\mathbf{r}) = \sum_{i=0}^{\infty} f(\epsilon_i - \mu) |\Psi_i|^2. \quad (3)$$

The arguments of $v[\rho, \mathbf{r}]$ are enclosed in square brackets to distinguish it from the output potential, which is the sum of the Hartree and exchange-correlation potentials given by

$$v(\rho, \mathbf{r}) = \delta(U + E_{xc}) / \delta\rho(\mathbf{r}). \quad (4)$$

In general, E_{xc} is temperature dependent; however, the dependence is very weak⁴ and will not be considered here. As expressed in Eq. (1), F is stationary with respect to $v[\rho, \mathbf{r}]$. Thus if F is evaluated with $v[\rho, \mathbf{r}]$ replaced by a reasonable approximation, the resulting error in F will be small. If $v[\rho, \mathbf{r}]$ is replaced by a functional such as $v(\rho, \mathbf{r})$ that satisfies $v(\rho, \mathbf{r})|_{\rho_{scf}} = v[\rho, \mathbf{r}]|_{\rho_{scf}}$ [ρ_{scf} is the fully converged local-density-approximation (LDA) electron density], then the resulting expression can be viewed as a new stationary functional,

$$F_H[\rho] = \int_{-\infty}^{\infty} d\epsilon \epsilon f(\epsilon - \mu) n[\epsilon, v(\rho, \mathbf{r})] - \int d\mathbf{r} \rho v(\rho, \mathbf{r}) + U[\rho] + E_{xc}[\rho] + \mu \left(\sum_i Z_i - \int_{-\infty}^{\infty} d\epsilon f(\epsilon - \mu) n[\epsilon, v(\rho, \mathbf{r})] \right) + k_B T \int_{-\infty}^{\infty} d\epsilon n[\epsilon, v(\rho, \mathbf{r})] \{ f(\epsilon - \mu) \ln f(\epsilon - \mu) + [1 - f(\epsilon - \mu)] \ln [1 - f(\epsilon - \mu)] \}. \quad (5)$$

Equations (4) and (5), which define the generalization of the Harris-Foulkes^{5,6} functional to finite T , are valid for arbitrary-shaped electron densities and potentials. However, we used the muffin-tin form of U and E_{xc} for testing purposes. Clearly $F_H[\rho] = F[\rho]$ if $\rho = \rho_{scf}$.

We now turn to the use of a locally self-consistent electron density, for the purpose of exploiting the stationarity of F . Specifically, we propose self-consistent calculations in which the Poisson equation is solved for the whole system, while the electron density is obtained by solving the multiple-scattering problem for clusters of finite size centered on each of the sites in turn. Thus, the total electron density $\rho_M(\mathbf{r})$ used in solving the Poisson equation is given by the sum of individual site densities, $\rho_M(\mathbf{r}) = \sum_i \rho_M^i(\mathbf{r}) \omega^i(\mathbf{r})$, where $\rho_M^i(\mathbf{r})$ corresponds to using an M -site cluster⁷ about site i and $\omega^i(\mathbf{r}) = 1$ (0) if \mathbf{r} is inside (outside) the i th Wigner-Seitz cell. We refer to $\rho_M(\mathbf{r})$ as the locally self-consistent density; the corresponding chemical potential is μ_M . The corresponding combination of Hartree and exchange-correlation potentials resulting from $\rho_M(\mathbf{r})$ is referred to as $v_M(\mathbf{r})$. Clearly, $\rho_M(\mathbf{r})$ approaches the true self-consistent density $\rho_{scf}(\mathbf{r})$, as M goes to infinity, i.e., $\rho_{\infty}(\mathbf{r}) = \rho_{scf}(\mathbf{r})$. In Fig. 1 we show, as a function of the cluster size used to determine $\rho_M(\mathbf{r})$, the error in the energy

obtained by using $\rho_M(\mathbf{r})$ and $v_M(\mathbf{r})$ in Eq. (5). Recall, we use the standard KKR method to evaluate the one-electron sum and chemical potential in Eq. (5). Interestingly, neglect of all but a single scatterer in solving the Schrödinger equation results in an error of only 1 mRy. Including a single shell of scatterers, $M = 13$, reduces this error to < 0.1 mRy, sufficient for all practical purposes.

We now discuss the Fermi function and approximations to it. The Fermi function f_F ,

$$f_F(z - \mu) = 1 / (e^{(z - \mu)\beta} + 1), \quad (6)$$

has a residue of $-k_B T$ at each of its Matsubara poles z_j^F , which are strung out like knots on a line passing through the real axis at μ ,

$$z_j^F = \mu + i(2j - 1)\pi k_B T \text{ for } j = 0, \pm 1, \pm 2, \dots \quad (7)$$

Thus, at finite T , contributions to the electron density and eigenvalue sum come from only these poles. Unfortunately, the poles are infinite in number and truncation at a finite number of poles is equivalent to using an approximate occupation function given on the real axis ($z = \epsilon$) by

$$f(\epsilon - \mu) \approx 1/2 + \beta(\epsilon - \mu) \sum_{j=1}^{j_{\max}} \frac{1}{[\beta(\epsilon - \mu)]^2 + [\pi(2j - 1)]^2}, \quad (8)$$

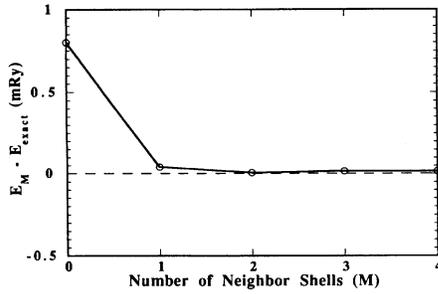


FIG. 1. Error in the energy of fcc Cu at $T=0$ K as a function of the cluster size used to determine $\rho_M(\mathbf{r})$.

where $\beta=1/k_B T$. As can be seen from Fig. 2, this approximation converges slowly in the number of poles, and is deficient because $f(\epsilon-\mu)\rightarrow 1/2$ as $\epsilon\rightarrow\pm\infty$. Thus, attempts to reduce computational effort by converting energy integrations to sums over Matsubara poles are fraught with difficulty.

Once again we turn to a stationary property of F . Observe that $f(\epsilon-\mu)$ appears implicitly, through $\rho(\mathbf{r})$, and explicitly in the eigenvalue sum and entropy. Only the explicit variation need be considered because the variation with respect to $\rho(\mathbf{r})$ vanishes. The remaining variation with respect to the explicit dependence on $f(\epsilon-\mu)$ vanishes^{8,9} provided $f(\epsilon-\mu)$ is given by the Fermi function. However, because of the stationarity of F with respect to $f(\epsilon-\mu)$ we have the liberty to modify $f(\epsilon-\mu)$ while doing minimal damage. If, instead of cutting the line of Matsubara poles at j_{\max} , we wrap it around a windlass of radius σ , chosen such that the poles on successive windings coincide, the following approximation is generated:

$$f_P(z-\mu) = \{[(z-\mu+\sigma)/\sigma]^{2P} + 1\}^{-1}. \quad (9)$$

The function $f_P(\epsilon-\mu)$, which we refer to as a band-Fermi function, is displayed in Fig. 2. The number of poles P is inversely related to T by $\beta=2P/\sigma$. Note that for $\mu-2\sigma\leq\epsilon$, $f_P(\epsilon-\mu)$ is a remarkably reliable approximation to $f_F(\epsilon-\mu)$. An example of the accuracy of $f_P(\epsilon-\mu)$ is shown in Fig. 3, where a comparison is made with the exact entropy $\pi k_B^2 T/3$ for the case of a constant unit density of states. In terms of calculating F , the fact that $f_P(\epsilon-\mu)$ is a very poor approximation below about

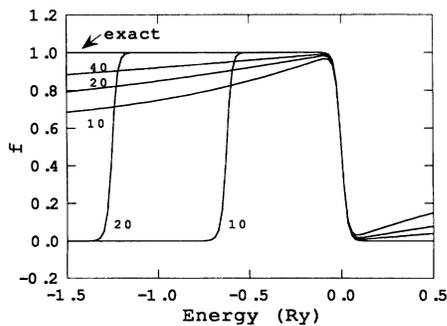


FIG. 2. The exact Fermi function at 2466 K is compared to values obtained from Eqs. (7) and (9) evaluated when a finite number of poles in the upper half-plane are included. The curves are labeled by the number of poles used. Those curves which drop to zero at both positive and negative energy are calculated with Eq. (9).

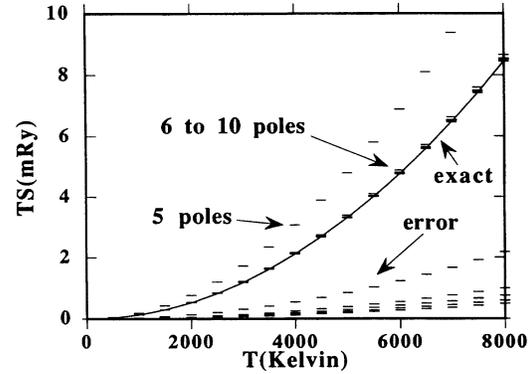


FIG. 3. Entropy terms $TS(f_P)$ and $TS(f_F)$ plotted as a function of T for the case of constant unit density of states. Also shown (lower points) is the error $T[S(f_P)-S(f_F)]$ magnified by a factor of 10 for $P=6, 7, 8, 9$, and 10.

$\mu-2\sigma$ is of no consequence, provided there are no eigenvalues within this energy interval. We can then use $f_P(\epsilon-\mu)$ for the valence band, choosing σ such that $f_P(\epsilon-\mu)$ goes to zero below the band and treat the remaining states as fully occupied unbanded core states.

Within multiple-scattering theory, occupying states according to $f_P(\epsilon-\mu)$ rather than $f_F(\epsilon-\mu)$ has two advantages. First, as previously suggested, the number of energies at which the Schrödinger equation needs to be solved is greatly reduced; there are only P poles in the upper half-plane. Second, all of these poles are removed from the real axis. An imaginary part in the energy introduces a damping into the real-space free particle propagator, making real-space multiple-scattering methods more rapidly convergent. For example, at $T=2446$ K ($k_B T=1/64$) and $\sigma=0.5$ Ry (2σ can span an occupied bandwidth of 1 Ry, which is typical of transition metals), $P=16$ and the pole closest to the real axis has an imaginary part of 49 mRy.

It is clear that the higher the electron temperature the simpler and more rapidly convergent real-space multiple-scattering methods utilizing $f_P(\epsilon-\mu)$ become. This prompts us to ask if it is possible to perform calculations at an elevated electron temperature and still obtain useful results for temperatures of physical interest, e.g., $T=0$ K. The answer is affirmative, because calculation of F provides an additional piece of information, the entropy. The negative of the entropy is the derivative of the free energy with respect to T . Additionally, we know that the entropy is zero at $T=0$ K (F is stationary with respect to T at $T=0$ K), is monotonically increasing, and has only odd derivatives at $T=0$ K. Thus, if we perform a calculation at an elevated T the free energy is known at that T , its derivative is known at two temperatures, and its third derivative as well as other odd derivatives are zero at $T=0$ K. Consequently, if the temperature at which the calculation is performed, T_f , is close to the temperature of interest then quadratic terms will dominate, and we can use

$$F(T) = E(T_f) - [T + (T-T_f)^2/2T_f]S(T_f) \quad (10)$$

to fit the value of $F(T_f)$ and its known derivatives. If the $T=0$ K energy is sought, it is convenient to use the fact that Eq. (10) implies $2F(T=0) = E(T_f) + F(T_f)$.¹⁰

In Fig. 4 we show the calculated free energy of Cu as a function of T together with a curve based on Eq. (10) using

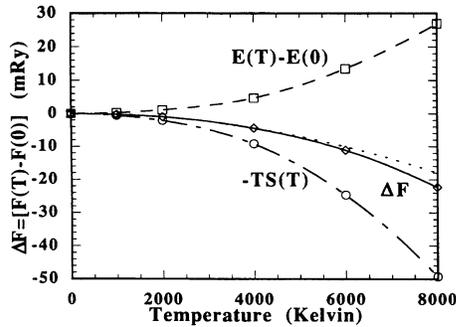


FIG. 4. Temperature dependence of the fully converged F , E , and $-TS$ for Cu. The calculations were performed using $f_P(\epsilon - \mu)$ with P taken to convergence and conventional reciprocal space KKR methods in both the charge self-consistency and total energy steps. The free energy based on Eq. (10) with $T_f = 4000$ K is shown by dots.

$T_f = 4000$ K. Determination of $E(T=0)$, using Eq. (10) and values of $T_f = 8000, 6000, 4000, 2000,$ and 1000 K, leads to errors of 2.30, 1.23, 0.15, 0.08, and 0.02 mRy, respectively. Thus, even if we desire the ground state energy, we can perform calculations at temperatures as high as 4000 K and still obtain useful precision.

In LDA bulk calculations the chemical potential is usually updated, along with ρ , at each iteration to maintain charge neutrality. This avoids the discomfoting situation of having an infinite Coulomb energy. Here, we suggest that it is possible to calculate μ_M and $\rho_M(\mathbf{r})$ in this manner, but to evaluate F_H without further adjustment of μ . Unfortunately, this means that the integral of the density of states is not ensured to balance the nuclear charge. The electron density $\rho_M(\mathbf{r})$ does provide a neutral charge density for evaluating the double counting terms in the Harris-Foulkes functional; however, occupation of the eigenstates up to μ_M gives the wrong number of electrons. Does miscounting the energy levels discredit the free energy? Presumably, μ_M is very close to the self-consistent value μ_{scf} . Since, in Eq. (5), μ appears explicitly only as the prefactor of one term in F_H and appears implicitly in variables with respect to which F_H is stationary, it follows that

$$\left. \frac{dF_H}{d\mu} \right|_{\mu_{scf}} = \sum_i Z_i - \int_{-\infty}^{\infty} d\epsilon f(\epsilon - \mu_{scf}) n(\epsilon) = 0. \quad (11)$$

In Fig. 5 we show the variation of F_H with respect to the deviation of μ from μ_{scf} . Also indicated are specific values of $\mu_{scf} - \mu_M$ for several choices of M . Clearly, an accuracy of tenths of mRy's is achieved for μ_M based on small clus-

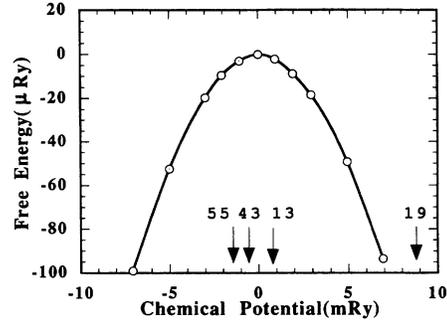


FIG. 5. Dependence on the chemical potential of the $T=0$ K free energy of Cu based on Eq. (5). The values of chemical potential used in evaluating Eq. (5) were displaced from the neutralizing value by small amounts. The electron density was taken to be $\rho_{\infty}(\mathbf{r})$. The difference between μ_{scf} and μ_M is indicated for $M = 13, 19, 43,$ and 55 .

ters. The important point is that, by exploiting the second-order variation with respect to μ , the calculations of the Green function necessary to readjust μ to its correct value are avoided.

Summarizing, we have pointed out that, at least for the self-consistency step, the $O(N^3)$ problem can be trivially circumvented because the values of $\rho_M(\mathbf{r})$ and μ_M emerging from locally self-consistent calculations are close enough to their exact values that they give free energies accurate to tenths of mRy. This local self-consistency algorithm is manifestly linear in the number of independent atoms in the system.

Additionally, we have shown that the band-Fermi function provides an elegant and reliable method of calculating the free energy at finite T . The use of $f_P(\epsilon - \mu)$ complements real-space multiple-scattering theory by limiting calculation to a small number of complex energies. The computational advantages of this formulation increase at elevated T because the number of poles decreases and the size of the required real-space cluster can be reduced. Even though we are usually interested in phenomenon at modest temperatures, we do not have to abandon the improved computational environment at elevated T because the entropy provides sufficient information to reliably fit the temperature dependence of the free energy.

This work was supported by the Division of Materials Science (G.M.S.) and by the Applied Mathematical Sciences Research Program (G.M.S., D.M.C.N., Y.W., W.A.S.), Office of Energy Research, U.S. DOE under Subcontract No. DEAC05-84OR21400 with Martin-Marietta Energy Systems, Inc. We would also like to acknowledge support from NATO (G.M.S., Z.S., W.M.T.).

¹D. J. Singh *et al.*, Phys. Rev. B **46**, 13 065 (1993).

²N. D. Mermin, Phys. Rev. **137**, A1441 (1965).

³U. Gupta and A. K. Rajagopal, Phys. Rep. **87**, 256 (1982).

⁴D. G. Kanhere *et al.*, Phys. Rev. B **33**, 490 (1986).

⁵J. Harris, Phys. Rev. B **31**, 1770 (1985).

⁶W. M. C. Foulkes and R. Haydock, Phys. Rev. B **39**, 12 520 (1985).

⁷G. Ries and H. Winter, J. Phys. F **9**, 1589 (1979).

⁸M. Weinert and J. W. Davenport, Phys. Rev. B **45**, 13 709 (1992).

⁹A. K. McMahan and M. Ross, Phys. Rev. B **15**, 718 (1977).

¹⁰M. J. Gillan, J. Phys. Condens. Matter **1**, 689 (1989).