

Multipole Representation of the Fermi Operator with Application to the Electronic Structure Analysis of Metallic Systems

Lin Lin and Jianfeng Lu

Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544

Roberto Car

Department of Chemistry and Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544

Weinan E

Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544

We propose a multipole representation of the Fermi-Dirac function and the Fermi operator, and use this representation to develop algorithms for electronic structure analysis of metallic systems. The new algorithm is quite simple and efficient. Its computational cost scales logarithmically with $\beta\Delta\epsilon$ where β is the inverse temperature, and $\Delta\epsilon$ is the width of the spectrum of the discretized Hamiltonian matrix.

PACS numbers:

I. INTRODUCTION

The Fermi operator, *i.e.* the Fermi-Dirac function of the system Hamiltonian, is a fundamental quantity in the quantum mechanics of many-electron systems and is ubiquitous in condensed matter physics. In the last decade the development of accurate and numerically efficient representations of the Fermi operator has attracted lot of attention in the quest for linear scaling electronic structure methods based on effective one-electron Hamiltonians. These approaches have numerical cost that scales linearly with N , the number of electrons, and thus hold the promise of making quantum mechanical calculations of large systems feasible. Achieving linear scaling in realistic calculations is very challenging. Formulations based on the Fermi operator are appealing because this operator gives directly the single particle density matrix without the need for Hamiltonian diagonalization. At finite temperature the density matrix can be expanded in terms of finite powers of the Hamiltonian, requiring computations that scale linearly with N owing to the sparse character of the Hamiltonian matrix[18]. These properties of the Fermi operator are valid not only for insulators but also for metals, making formulations based on the Fermi operator particularly attractive.

Electronic structure algorithms using a Fermi operator expansion (FOE) were introduced by Baroni and Giannozzi [1] and Goedecker and co-workers [2, 3] (see also the review article [4]). These authors proposed polynomial and rational approximations of the Fermi operator. Major improvements were made recently in a series of publications by Parrinello and co-authors [5, 6, 7, 8, 9, 10], in which a new form of Fermi operator expansion was introduced based on the grand canonical formalism.

From the viewpoint of efficiency, a major concern is the cost of the representations of the Fermi operator as a function of $\beta\Delta\epsilon$ where β is the inverse temperature and $\Delta\epsilon$ is the spectral width of the Hamiltonian matrix. The cost of the original FOE proposed by Goedecker *et al* scales as $\beta\Delta\epsilon$. The fast polynomial summation technique introduced by Head-Gordon *et al*[11, 12] reduces the cost to $(\beta\Delta\epsilon)^{1/2}$. The cost of the hybrid algorithm proposed by Parrinello *et al* in a recent preprint [10] scales as $(\beta\Delta\epsilon)^{1/3}$.

The main purpose of this paper is to present a strategy that reduces the cost to logarithmic scaling $(\ln\beta\Delta\epsilon) \cdot (\ln\ln\beta\Delta\epsilon)$, thus greatly improving the efficiency and accuracy of numerical FOEs. Our approach is based on the exact pole expansion of the Fermi-Dirac function which underlies the Matsubara formalism of finite temperature Green's functions in many-body physics [13]. It is natural to consider a multipole representation of this expansion to achieve better efficiency, as was done in the fast multipole method [14]. Indeed, as we will show below, the multipole expansion that we propose does achieve logarithmic scaling. We believe that this representation will be quite useful both as a theoretical tool and as a starting point for computations. As an application of the new formalism, we present an algorithm for electronic structure calculation that has the potential to become an efficient linear scaling algorithm for metallic systems.

The remaining of the paper is organized as follows. In the next section, we introduce the multipole representation of Fermi operator. In Section III, we present the FOE algorithm based on the multipole representation and analyze its computational cost. Three examples illustrating the algorithm are discussed in Section IV. We conclude the paper with some remarks on future directions.

II. MULTIPOLE REPRESENTATION FOR THE FERMI OPERATOR

Given the effective one-particle Hamiltonian \mathbf{H} , the inverse temperature $\beta = 1/k_B T$ and the chemical potential μ , the finite temperature single-particle density matrix of the system is given by the Fermi operator

$$\rho = \frac{2}{1 + \exp(\beta(\mathbf{H} - \mu))} = 1 - \tanh\left(\frac{\beta}{2}(\mathbf{H} - \mu)\right), \quad (1)$$

where \tanh is the hyperbolic tangent function. The Matsubara representation of the Fermi-Dirac function is given by

$$\rho = 1 - 4\text{Re} \sum_{l=1}^{\infty} \frac{1}{\beta(\mathbf{H} - \mu) - (2l-1)\pi i}. \quad (2)$$

This representation originates from the pole expansion (see, for example, [15, 16]) of the meromorphic function \tanh

$$\tanh z = \sum_{l=-\infty}^{\infty} \frac{1}{z - \frac{2l-1}{2}\pi i}. \quad (3)$$

In particular, for x real (which is the case when \mathbf{H} is self-adjoint), we have

$$\tanh x = 4\text{Re} \sum_{l=1}^{\infty} \frac{1}{2x - (2l-1)\pi i}. \quad (4)$$

To make the paper self-contained, we provide in Appendix A a simple derivation of this representation. It should be emphasized that (2) is *exact*. We notice that the expansion (2) can also be understood as the $P \rightarrow \infty$ limit of an exact Fermi operator expansion proposed by Parrinello and co-authors in [5, 6, 7, 8, 9, 10].

The summation in (2) can be seen as a summation of residues contributed from the poles $\{(2l-1)\pi i\}$, with l a positive integer, on the imaginary axis. This suggests to look for a multipole expansion of the contributions from the poles, as done in the fast multipole method (FMM) [14]. To do so, we use a dyadic decomposition of the poles, in which the n -th group contains terms from $l = 2^{n-1}$ to $l = 2^n - 1$, for a total of 2^{n-1} terms (see Figure 1 for illustration). We decompose the summation in Eq.(2) accordingly, denoting $x = \beta(\mathbf{H} - \mu)$ for simplicity

$$\sum_{l=1}^{\infty} \frac{1}{x - (2l-1)\pi i} = \sum_{n=1}^{\infty} \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x - (2l-1)\pi i} = \sum_{n=1}^{\infty} S_n. \quad (5)$$

The basic idea is to combine the simple poles into a set of multipoles at $l = l_n$, where l_n is taken as the midpoint of the interval $[2^{n-1}, 2^n - 1]$

$$l_n = \frac{3 \cdot 2^{n-1} - 1}{2}. \quad (6)$$

Then the S_n term in the above equation can be written as

$$\begin{aligned} S_n &= \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x - (2l-1)\pi i - 2(l-l_n)\pi i} \\ &= \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x - (2l_n-1)\pi i} \sum_{\nu=0}^{\infty} \left(\frac{2(l-l_n)\pi i}{x - (2l_n-1)\pi i} \right)^{\nu} \\ &= \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x - (2l_n-1)\pi i} \sum_{\nu=0}^{P-1} \left(\frac{2(l-l_n)\pi i}{x - (2l_n-1)\pi i} \right)^{\nu} \\ &\quad + \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x - (2l-1)\pi i} \left(\frac{2(l-l_n)\pi i}{x - (2l_n-1)\pi i} \right)^P \end{aligned} \quad (7)$$

In deriving Eq. (7) we used the result for the sum of a geometric series. Using the fact that x is real, the second term in Eq. (7) can be bounded by

$$\sum_{l=2^{n-1}}^{2^n-1} \left| \frac{1}{x - (2l-1)\pi i} \right| \left| \frac{2(l-l_n)\pi i}{x - (2l_n-1)\pi i} \right|^P \leq \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{|(2l-1)\pi|} \left| \frac{2(l-l_n)}{2l_n-1} \right|^P \leq \frac{1}{2\pi} \frac{1}{3^P} \quad (8)$$

Therefore, we can approximate the sum S_n by the first P terms, and the error decays exponentially with P :

$$\left| S_n(x) - \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x - (2l_n - 1)\pi i} \sum_{\nu=0}^{P-1} \left(\frac{2(l - l_n)\pi i}{x - (2l_n - 1)\pi i} \right)^\nu \right| \leq \frac{1}{2\pi} \frac{1}{3^P}, \quad (9)$$

uniformly in x . The above analysis is of course standard from the view point of the fast multipole method [14]. The overall philosophy is also similar: given a preset error tolerance, one selects P , the number of terms to retain in S_n , according to Eq. (9).

Interestingly, the remainder of the sum in Eq. (2) from $l = m$ to ∞ has an explicit expression

$$\text{Re} \sum_{l=m}^{\infty} \frac{1}{2x - (2l - 1)i\pi} = \frac{1}{2\pi} \text{Im} \psi \left(m - \frac{1}{2} + \frac{i}{\pi} x \right), \quad (10)$$

where ψ is the digamma function $\psi(z) = \Gamma'(z)/\Gamma(z)$. It is well known [15] that the digamma function has the following asymptotic expansion

$$\psi(z) \sim \ln(z) - \frac{1}{2z} - \frac{1}{12z^2} + \mathcal{O}\left(\frac{1}{z^4}\right), \quad |\arg z| \leq \pi \text{ and } |z| \rightarrow \infty. \quad (11)$$

Therefore,

$$\begin{aligned} \text{Im} \psi \left(m - \frac{1}{2} + \frac{i}{\pi} x \right) &\sim \text{Im} \ln \left(m - \frac{1}{2} - \frac{i}{\pi} x \right) + \mathcal{O}\left(\frac{1}{m^2}\right) \\ &= \arctan \left(\frac{2x}{(2m - 1)\pi} \right) + \mathcal{O}\left(\frac{1}{m^2}\right), \quad m \rightarrow \infty. \end{aligned} \quad (12)$$

Figure 2 shows that the asymptotic approximation (12) is already rather accurate when $m = 10$.

Eq. (12) also shows the effectiveness of the multipole representation from the viewpoint of traditional polynomial approximations. At zero temperature, the Fermi-Dirac function is a step function that cannot be accurately approximated by any finite order polynomial. At finite but low temperature, it is a continuous function with a very large derivative at $x = 0$, *i.e.* when the energy equals the chemical potential μ . The magnitude of this derivative becomes smaller and, correspondingly, the Fermi function becomes smoother as the temperature is raised. One can use the value of the derivative of the Fermi function at $x = 0$ to measure the difficulty of an FOE. After eliminating the first m terms in the expansion, Eq. (12) shows that asymptotically the derivative is multiplied by the factor $\frac{2}{(2m-1)\pi}$, which is equivalent to a rescaling of the temperature by the same factor. In particular, if we explicitly include the first 2^N terms in the multipole representation of the Fermi operator, we are left with a remainder which is well approximated by Eq. (12), so that, effectively, the difficulty is reduced by a factor 2^N . As a matter of fact standard polynomials approximations, such as the Chebyshev expansion, can be used to efficiently represent the remainder in Eq. (10) even at very low temperature.

In summary, we arrive at the following multipole representation for the Fermi operator

$$\begin{aligned} \rho = 1 - 4\text{Re} \sum_{n=1}^N \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{\beta(\mathbf{H} - \mu) - (2l_n - 1)\pi i} \sum_{\nu=0}^{P-1} \left(\frac{2(l - l_n)\pi i}{\beta(\mathbf{H} - \mu) - (2l_n - 1)\pi i} \right)^\nu \\ - \frac{2}{\pi} \text{Im} \psi \left(2^N - \frac{1}{2} + \frac{i}{2\pi} \beta(\mathbf{H} - \mu) \right) + \mathcal{O}(N/3^P). \end{aligned} \quad (13)$$

The multipole part is evaluated directly as discussed below, and the remainder is evaluated with the standard polynomial method.

III. NUMERICAL CALCULATION AND ERROR ANALYSIS

To show the power of the multipole expansion, we discuss a possible algorithm to compute the Fermi operator in electronic structure calculations and present a detailed analysis of its cost in terms of $\beta\Delta\epsilon$. Given the Hamiltonian matrix \mathbf{H} , it is straightforward to compute the density matrix ρ from the multipole expansion if we can calculate the Green's functions $\mathbf{B}_{l_n} = [\beta(\mathbf{H} - \mu) - (2l_n - 1)\pi i]^{-1}$ for different n .

A possible way to calculate the inverse matrices is by the Newton-Schulz iteration. For any non-degenerate matrix \mathbf{A} , the Newton-Schulz iteration computes the inverse $\mathbf{B} = \mathbf{A}^{-1}$ as

$$\mathbf{B}_{k+1} = 2\mathbf{B}_k - \mathbf{B}_k\mathbf{A}\mathbf{B}_k. \quad (14)$$

The iteration error is measured by the spectral radius, *i.e.* the eigenvalue of largest magnitude, of the matrix $\mathbf{I} - \mathbf{B}\mathbf{B}_k$ where \mathbf{I} is the identity matrix. In the following we denote the spectral radius of the matrix \mathbf{A} by $\sigma(\mathbf{A})$. Then the spectral radius at the k -th step of the Newton-Schulz iteration is $\mathbf{R}_k = \sigma(\mathbf{I} - \mathbf{B}\mathbf{B}_k)$ and

$$\sigma(\mathbf{R}_{k+1}) = \sigma(\mathbf{R}_k)^2 = \sigma(\mathbf{R}_0)^{2^{k+1}}. \quad (15)$$

Thus the Newton-Schulz iteration has quadratic convergence. With a proper choice of the initial guess (see [9]), the number of iterations required to converge is bounded by a constant, and this constant depends only on the target accuracy.

The remainder, *i.e.* the term associated to the digamma function in Eq. (13), can be evaluated by standard polynomial approximations such as the Chebyshev expansion. The order of Chebyshev polynomials needed for a given target accuracy is proportional to $\beta\Delta\epsilon/2^{N+1}$ (see [17, Appendix]).

Except for the error coming from the truncated multipole representation, the main source of error in applications comes from the numerical approximation of the Green's functions \mathbf{B}_{l_n} . To understand the impact of this numerical error on the representation of the Fermi operator, let us rewrite

$$\mathbf{S}_n = \sum_{l=2^{n-1}}^{2^n-1} \mathbf{B}_{l_n} \sum_{\nu=0}^{P-1} (-2(l-l_n)\pi i \mathbf{B}_{l_n})^\nu = \sum_{\nu=0}^{P-1} \mathbf{B}_{l_n}^{\nu+1} \sum_{l=2^{n-1}}^{2^n-1} (-2(l-l_n)\pi i)^\nu.$$

The factor $\sum_l (-2(l-l_n)\pi i)^\nu$ is large, but we can control the total error in S_n in terms of the spectral radius $\sigma(\mathbf{B}_{l_n} - \widehat{\mathbf{B}}_{l_n})$. Here $\widehat{\mathbf{B}}_{l_n}$ is the numerical estimate of \mathbf{B}_{l_n} .

The error is bounded by

$$\sigma(\widehat{\mathbf{S}}_n - \mathbf{S}_n) \leq \sum_{\nu=0}^{P-1} 2^{n-1} (2^{n-1}\pi)^\nu \sigma(\mathbf{B}^{\nu+1} - \widehat{\mathbf{B}}^{\nu+1}) \leq \sum_{\nu=0}^{P-1} (2^{n-1}\pi)^{\nu+1} \sigma(\mathbf{B}^{\nu+1} - \widehat{\mathbf{B}}^{\nu+1}), \quad (16)$$

where we have omitted the subscript l_n in \mathbf{B}_{l_n} and in $\widehat{\mathbf{B}}_{l_n}$. In what follows the quantity $\sum_{\nu=0}^{P-1} (2^{n-1}\pi)^{\nu+1} \sigma(\mathbf{B}^{\nu+1} - \widehat{\mathbf{B}}^{\nu+1})$ will be denoted by e_P . Then we have

$$\begin{aligned} e_P &= \sum_{\nu=0}^{P-1} (2^{n-1}\pi)^{\nu+1} \sigma((\mathbf{B}^\nu - \widehat{\mathbf{B}}^\nu)\mathbf{B} + (\widehat{\mathbf{B}}^\nu - \mathbf{B}^\nu)(\mathbf{B} - \widehat{\mathbf{B}}) + \mathbf{B}^\nu(\mathbf{B} - \widehat{\mathbf{B}})) \\ &\leq \sum_{\nu=1}^{P-1} (2^{n-1}\pi)^{\nu+1} (\sigma(\mathbf{B}) + \sigma(\mathbf{B} - \widehat{\mathbf{B}})) \sigma(\mathbf{B}^\nu - \widehat{\mathbf{B}}^\nu) + \sum_{\nu=0}^{P-1} (2^{n-1}\pi)^{\nu+1} \sigma(\mathbf{B})^\nu \sigma(\mathbf{B} - \widehat{\mathbf{B}}). \end{aligned} \quad (17)$$

Here we took into account the fact that the $\nu = 0$ term in the first summation is equal to zero, and have used the properties $\sigma(\mathbf{A} + \mathbf{B}) \leq \sigma(\mathbf{A}) + \sigma(\mathbf{B})$, and $\sigma(\mathbf{A}\mathbf{B}) \leq \sigma(\mathbf{A})\sigma(\mathbf{B})$, respectively.

Noting that $2^{n-1}\pi\sigma(\mathbf{B}_{l_n}) \leq 1/3$ and changing ν to $\nu + 1$ in the first summation, we can rewrite e_P as

$$\begin{aligned} e_P &\leq \left(\frac{1}{3} + 2^{n-1}\pi\sigma(\mathbf{B} - \widehat{\mathbf{B}})\right) \sum_{\nu=0}^{P-2} (2^{n-1}\pi)^{\nu+1} \sigma(\mathbf{B}^{\nu+1} - \widehat{\mathbf{B}}^{\nu+1}) + \sum_{\nu=0}^{P-1} \frac{1}{3^\nu} 2^{n-1}\pi\sigma(\mathbf{B} - \widehat{\mathbf{B}}) \\ &\leq \left(\frac{1}{3} + 2^{n-1}\pi\sigma(\mathbf{B} - \widehat{\mathbf{B}})\right) e_{P-1} + \frac{3}{2} 2^{n-1}\pi\sigma(\mathbf{B} - \widehat{\mathbf{B}}) \\ &= \left(\frac{1}{3} + e_1\right) e_{P-1} + \frac{3}{2} e_1. \end{aligned} \quad (18)$$

In the last equality, we used the fact that $e_1 = 2^{n-1}\pi\sigma(\mathbf{B} - \widehat{\mathbf{B}})$. Therefore, the error e_P satisfies the following recursion formula

$$e_P + \frac{3e_1/2}{e_1 - 2/3} \leq \left(\frac{1}{3} + e_1\right) \left(e_1 + \frac{3e_1/2}{e_{P-1} - 2/3}\right) \leq \left(\frac{1}{3} + e_1\right)^{P-1} \left(e_1 + \frac{3e_1/2}{e_1 - 2/3}\right). \quad (19)$$

Taking $e_1 \leq \frac{2}{3}$, we have

$$e_P \leq e_1 = 2^{n-1} \pi \sigma(\mathbf{B} - \hat{\mathbf{B}}). \quad (20)$$

Therefore, using Eq. (15) we find that the number k of Newton-Schulz iterations must be bounded as dictated by the following inequality in order for the error $\sigma(\hat{\mathbf{S}}_n - \mathbf{S}_n)$ to be $\leq 10^{-D}/N$.

$$2^{n-1} \sigma(\mathbf{R}_0)^{2^k} \leq \frac{10^{-D}}{N}. \quad (21)$$

Here we have used the fact that $\sigma(\mathbf{B}_{l_n}) \leq 1/\pi$ for any n . Each Newton-Schulz iteration requires two matrix by matrix multiplications, and the number of matrix by matrix multiplications needed in the Newton-Schulz iteration for \mathbf{B}_{l_n} with $n < N$ is bounded by

$$2 \log_2 \left(\frac{D \log_2 10 + N + \log_2 N}{-\log_2 \sigma(\mathbf{R}_0)} \right). \quad (22)$$

To obtain a target accuracy $\sigma(\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}) \leq 10^{-D}$ for a numerical estimate $\hat{\boldsymbol{\rho}}$ of the density matrix, taking into account the operational cost of calculating the remainder and the direct multipole summation in the FOE, the number of matrix by matrix multiplications n_{MM} is bounded by

$$n_{\text{MM}} \leq 2N \log_2 N + C_1 N + C_2 2^{-N-1} \beta \Delta \epsilon. \quad (23)$$

Here we used the property: $\log_2(x+y) \leq \log_2 x + \log_2 y$ when $x, y \geq 2$, and defined the constant C_1 as follows:

$$C_1 = \frac{2}{N} \sum_{n=1}^N \log_2 \left(\frac{D \log_2 10 + \log_2 N}{-\log_2 \sigma((\mathbf{R}_0)_{l_n})} \right). \quad (24)$$

The dependence on $2^{-N-1} \beta \Delta \epsilon$ in the last term on the right hand side of (23) comes from Chebyshev expansion used to calculate the remainder. From numerical calculations on model systems, the constant C_1 and C_2 will be shown to be rather small. Finally, choosing $N \propto \ln(\beta \Delta \epsilon)$, we obtain

$$n_{\text{MM}} \propto (\ln \beta \Delta \epsilon) \cdot (\ln \ln \beta \Delta \epsilon) \quad (25)$$

with a small prefactor.

IV. NUMERICAL EXAMPLES

We illustrate the algorithm in three simple cases. The first is an off-lattice one dimensional model defined in a supercell with periodic boundary conditions. In this example, we discretize the Hamiltonian with the finite difference method, resulting in a very broad spectrum with a width of about 2000eV, and we choose a temperature as low as 32K. In the second example we consider a nearest neighbor tight binding Hamiltonian in a three dimensional simple cubic lattice and set the temperature to 100K. In the third example we consider a three dimensional Anderson model with random on-site energy on a simple cubic lattice at 100K.

A. One dimensional model with large spectral width

In this example, a one dimensional crystal is described by a periodic supercell with 10 atoms, evenly spaced. We take the distance between adjacent atoms to be $a = 5.29\text{\AA}$. The one-particle Hamiltonian is given by

$$\mathbf{H} = -\frac{1}{2} \frac{\partial^2}{\partial x^2} + V. \quad (26)$$

The potential V is given by a sum of Gaussians centered at the atoms with width $\sigma = 1.32\text{\AA}$ and depth $V_0 = 13.6\text{eV}$. The kinetic energy is discretized using a simple 3-point finite difference formula, resulting in a Hamiltonian \mathbf{H} with a discrete eigenvalue spectrum with lower and upper eigenvalues equal to $\epsilon_- = 6.76\text{eV}$ and $\epsilon_+ = 1959\text{eV}$, respectively. Various temperatures from 1024K to 32K were tried. Figure 3 reports the linear-log graph of n_{MM} , the number of

matrix by matrix multiplications needed to evaluate the density matrix using our FOE, versus $\beta\Delta\epsilon$, with $\beta\Delta\epsilon$ plotted in a logarithmic scale. The logarithmic dependence can be clearly seen. The prefactor of the logarithmic dependence is rather small: when $\beta\Delta\epsilon$ is doubled, a number of additional matrix multiplications equal to 17 is required to achieve two-digit accuracy ($D = 2$), a number equal to 19 is needed for $D = 4$, and a number equal to 21 is needed for $D = 6$, respectively. The observed D -dependence of the number of matrix multiplications agrees well with the prediction in (23).

In order to assess the validity of the criterion for the number of matrix multiplications given in Eq. (23), we report in Table I the calculated relative energy error and relative density error, respectively, at different temperatures, when the number of matrix multiplications is bounded as in formula (23) using different values for D . The relative energy error, $\Delta\epsilon_{\text{rel}}$, measures the accuracy in the calculation of the total electronic energy corresponding to the supercell $E = \text{Tr}(\rho\mathbf{H})$. It is defined as

$$\Delta\epsilon_{\text{rel}} = \frac{|\hat{E} - E|}{|E|}. \quad (27)$$

Similarly the relative L^1 error in the density function in real space is defined as

$$\Delta\rho_{\text{rel}} = \frac{\text{Tr}|\hat{\rho} - \rho|}{\text{Tr}\rho}. \quad (28)$$

Because $\text{Tr}\rho = N_e$, where N_e is the total number of electrons in the supercell, $\Delta\rho_{\text{rel}}$ is the same as the L^1 density error per electron. Table I shows that for all the values of $\beta\Delta\epsilon$, our algorithm gives a numerical accuracy that is even better than the target accuracy D . This is not surprising because our theoretical analysis was based on the most conservative error estimates.

B. Periodic three dimensional tight-binding model

In this example we consider a periodic three dimensional single-band tight-binding Hamiltonian in a simple cubic lattice. The Hamiltonian, which can be viewed as the discretized form of a free-particle Hamiltonian, is given in second quantized notation by:

$$\mathbf{H} = -t \sum_{\langle i,j \rangle} c_i^+ c_j, \quad (29)$$

where the sum includes the nearest neighbors only. Choosing a value of 2.27eV for the hopping parameter t the band extrema occur at $\epsilon_+ = 13.606\text{eV}$, and at $\epsilon_- = -13.606\text{eV}$, respectively. In the numerical calculation we consider a periodically repeated supercell with 1000 sites and chose a value of 100K for the temperature. Table II shows the dependence of n_{MM} , $\Delta\epsilon_{\text{rel}}$, and $\Delta\rho_{\text{rel}}$ on the chemical potential μ , for different D choices. Compared to the previous one dimensional example in which $\beta\Delta\epsilon$ was as large as 7.12×10^5 , here $\beta\Delta\epsilon = 1600$ due to the much smaller spectral width of the tight-binding Hamiltonian. When $\mu = 0$ the chemical potential lies exactly in the middle of the spectrum. This symmetry leads to a relative error as low as 10^{-19} for the density function.

C. Three dimensional disordered Anderson model

In this example we consider an Anderson model with on-site disorder on a simple cubic lattice. The Hamiltonian is given by

$$\mathbf{H} = -t \sum_{\langle i,j \rangle} c_i^+ c_j + \sum_i \epsilon_i c_i^+ c_i. \quad (30)$$

This Hamiltonian contains random on-site energies ϵ_i uniformly distributed in the interval $[-1.13\text{eV}, 1.13\text{eV}]$, and we use the same hopping parameter t as in the previous (ordered) example. In the numerical calculation we consider, as before, a supercell with 1000 sites with periodic boundary conditions, and choose again a temperature of 100K. In one realization of disorder corresponding to a particular set of random on-site energies, the spectrum has extrema at $\epsilon_+ = 13.619\text{eV}$ and at $\epsilon_- = -13.676\text{eV}$. The effect of disorder on the density function is remarkable: while in the periodic tight-binding case the density was uniform, having the same constant value at all the lattice sites, now the density is a random function in the lattice sites within the supercell. Table III reports for the disordered model the same data that were reported in Table II for the ordered model. We see that the accuracy of our numerical FOE is the same in the two cases, irrespective of disorder. The only difference is that the super convergence due to symmetry for $\mu = 0$ no longer exists in the disordered case.

V. CONCLUSION

We proposed a multipole representation for the Fermi operator. Based on this expansion, a rather simple and efficient algorithm was developed for electronic structure analysis. We have shown that the number of matrix by matrix multiplication that are needed scales as $(\ln \beta \Delta \epsilon) \cdot (\ln \ln \beta \Delta \epsilon)$ with very small overhead. Numerical examples show that the algorithm is promising and has the potential to be applied to metallic systems.

We have only considered the number of matrix matrix multiplications as a measure for the computational cost. The real operational count should of course take into account the cost of multiplying two matrices, and hence depends on how the matrices are represented. This is work in progress.

APPENDIX A: MITTAG-LEFFLER'S THEOREM AND POLE EXPANSION FOR HYPERBOLIC TANGENT FUNCTION

To obtain the pole expansion for hyperbolic tangent function $\tanh(z)$, we need a special case of the general Mittag-Leffler's theorem on the expansions of meromorphic functions (see, for example, [15, 16]).

Theorem 1 (Mittag-Leffler). *If a function $f(z)$ analytic at the origin has no singularities other than poles for finite z , and if we can choose a sequence of contours C_m about $z = 0$ tending to infinity, such that $|f(z)| \leq M$ on C_m and $\int_{C_m} |dz/z|$ is uniformly bounded, then we have*

$$f(z) = f(0) + \lim_{m \rightarrow \infty} \{P_m(z) - P_m(0)\}, \quad (\text{A1})$$

where $P_m(z)$ is the sum of the principal parts of $f(z)$ at all poles within C_m .

For $\tanh(z) = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)}$, it is analytic at the origin and $\tanh(0) = 0$. The function has simple poles at $z_l = (l - 1/2)\pi i$, $l \in \mathbb{Z}$ with principle parts $(z - z_l)^{-1}$. Let us take the contours as

$$C_m = \{x \pm im\pi \mid |x| \leq m\pi\} \cup \{\pm m\pi + iy \mid |y| \leq m\pi\}, \quad m \in \mathbb{Z}_+,$$

it is then easy to verify that C_m satisfy the conditions in Theorem 1. According to Theorem 1,

$$\tanh(z) = \tanh(0) + \lim_{m \rightarrow \infty} \sum_{l=-m+1}^m \left(\frac{1}{z - z_l} + \frac{1}{z_l} \right). \quad (\text{A2})$$

By symmetry of z_l , the second term within the brackets cancels, and we arrive at (3).

Acknowledgement: We thank M. Ceriotti and M. Parrinello for useful discussions. This work was partially supported by DOE under Contract No. DE-FG02-03ER25587 and by ONR under Contract No. N00014-01-1-0674 (W. E. L. L., and J. L.), and by DOE under Contract No. DE-FG02-05ER46201 and NSF-MRSEC Grant DMR-02B706 (R. C. and L. L.).

-
- [1] S. Baroni and P. Giannozzi, *Europhys. Lett.* **17**, 547 (1992).
 - [2] S. Goedecker and L. Colombo, *Phys. Rev. Lett.* **73**, 122 (1994).
 - [3] S. Goedecker and M. Teter, *Phys. Rev. B* **51**, 9455 (1995).
 - [4] S. Goedecker, *Rev. Mod. Phys.* **71**, 1085 (1999).
 - [5] F. Krajewski and M. Parrinello, *Phys. Rev. B* **71**, 233105 (2005).
 - [6] F. Krajewski and M. Parrinello, *Phys. Rev. B* **73**, 41105 (2006).
 - [7] F. Krajewski and M. Parrinello, *Phys. Rev. B* **74**, 125107 (2006).
 - [8] F. Krajewski and M. Parrinello, *Phys. Rev. B* **75**, 235108 (2007).
 - [9] M. Ceriotti, T. Kühne, and M. Parrinello, *J. Chem. Phys.* **129**, 024707 (2008).
 - [10] M. Ceriotti, T. Kühne, and M. Parrinello, arXiv:0809.2232v1 (2008).
 - [11] W. Liang, C. Saravanan, Y. Shao, R. Baer, A. T. Bell, and M. Head-Gordon, *J. Chem. Phys.* **119**, 4117 (2003).
 - [12] W. Liang, R. Baer, C. Saravanan, Y. Shao, A. T. Bell, and M. Head-Gordon, *J. Comput. Phys.* **194**, 575 (2004).
 - [13] G. Mahan, *Many-particle Physics* (Plenum Pub Corp, 2000).
 - [14] L. Greengard and V. Rokhlin, *J. Comput. Phys.* **73**, 325 (1987).
 - [15] H. Jeffreys and B. S. Jeffreys, *Methods of mathematical physics* (Cambridge, at the University Press, 1956), 3rd ed.

- [16] L. V. Ahlfors, *Complex analysis* (McGraw-Hill Book Co., New York, 1978), 3rd ed.
- [17] R. Baer and M. Head-Gordon, *J. Chem. Phys.* **107**, 10003 (1997).
- [18] When the effective Hamiltonian depends on the density, such as *e.g.* in density functional theory, the number of iterations needed to achieve self-consistency may be an additional source of size dependence. This issue has received little attention so far in the literature and will not be considered in this paper.

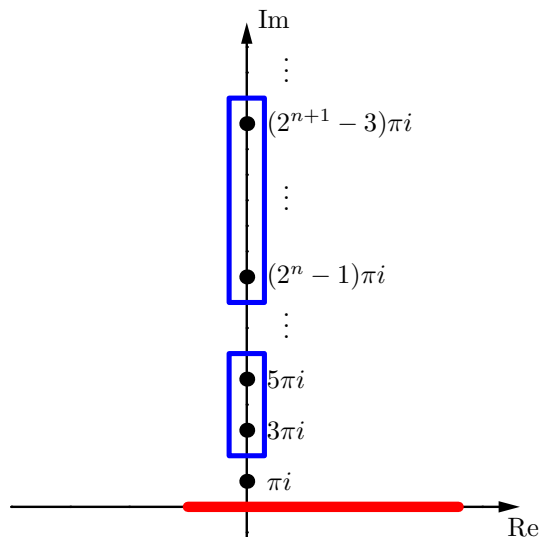


FIG. 1: (color online) Illustration of the pole decomposition (13). From 2^n to $2^{n+1} - 1$ poles are grouped together as shown in the figure. The spectrum is indicated by the red line on the real axis.

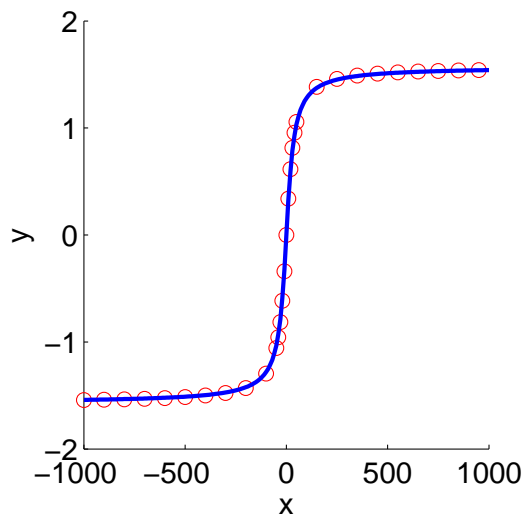


FIG. 2: (color online) The function $\text{Im } \psi\left(m - \frac{1}{2} + \frac{i}{\pi}x\right)$ (red circle), *i.e.* the remainder of the pole expansion in Eq. (13) is compared with the function $\arctan\left(\frac{2x}{(2^m-1)\pi}\right)$ (blue solid line) for $m = 10$

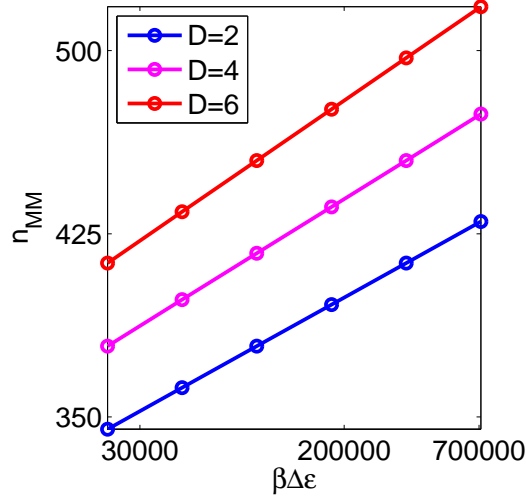


FIG. 3: (color online) Linear-log plot of the number of matrix matrix multiplications n_{MM} versus $\beta\Delta\epsilon$. n_{MM} depends logarithmically on $\beta\Delta\epsilon$ with a small constant prefactor.

T	$\beta\Delta\epsilon$	$\Delta\epsilon_{\text{rel}}$			$\Delta\rho_{\text{rel}}$		
		$D=2$	$D=4$	$D=6$	$D=2$	$D=4$	$D=6$
1024K	2.22×10^4	1.64×10^{-3}	5.98×10^{-6}	3.31×10^{-8}	4.21×10^{-4}	2.23×10^{-6}	1.50×10^{-8}
512K	4.44×10^4	1.73×10^{-3}	6.49×10^{-6}	3.70×10^{-8}	4.63×10^{-4}	2.52×10^{-6}	1.74×10^{-8}
256K	8.89×10^4	1.78×10^{-3}	6.83×10^{-6}	3.96×10^{-8}	4.77×10^{-4}	2.62×10^{-6}	1.81×10^{-8}
128K	1.78×10^5	1.74×10^{-3}	6.55×10^{-6}	3.75×10^{-8}	5.04×10^{-4}	2.80×10^{-6}	1.95×10^{-8}
64K	3.56×10^5	1.75×10^{-3}	6.62×10^{-6}	3.80×10^{-8}	4.92×10^{-4}	2.70×10^{-6}	1.86×10^{-8}
32K	7.12×10^5	1.76×10^{-3}	6.66×10^{-6}	3.82×10^{-8}	4.84×10^{-4}	2.64×10^{-6}	1.80×10^{-8}

TABLE I: One dimensional Hamiltonian model of Section IV A. Relative energy error $\Delta\epsilon_{\text{rel}}$ and relative L^1 density error $\Delta\rho_{\text{rel}}$ for a large range of values of $\beta\Delta\epsilon$ and several values of D .

μ	$D=4$			$D=8$		
	n_{MM}	$\Delta\epsilon_{\text{rel}}$	$\Delta\rho_{\text{rel}}$	n_{MM}	$\Delta\epsilon_{\text{rel}}$	$\Delta\rho_{\text{rel}}$
-10.88eV	320	4.09×10^{-9}	2.31×10^{-10}	376	2.27×10^{-13}	2.37×10^{-14}
-5.44eV	308	1.48×10^{-9}	3.15×10^{-11}	356	4.77×10^{-13}	2.52×10^{-15}
0.00eV	305	1.55×10^{-9}	6.26×10^{-19}	357	2.98×10^{-15}	6.26×10^{-19}
5.44eV	308	1.45×10^{-8}	1.34×10^{-12}	356	5.36×10^{-13}	1.07×10^{-16}
10.88eV	320	1.69×10^{-8}	1.78×10^{-13}	376	1.09×10^{-12}	1.80×10^{-17}

TABLE II: Three dimensional periodic tight binding model of Section IV B. Number of matrix matrix multiplications n_{MM} , relative energy error $\Delta\epsilon_{\text{rel}}$, and relative L^1 density error $\Delta\rho_{\text{rel}}$. For $\mu = 0$, the algorithm achieves machine accuracy for the absolute error of the density function as a consequence of symmetry.

μ	$D=4$			$D=8$		
	n_{MM}	$\Delta\epsilon_{\text{rel}}$	$\Delta\rho_{\text{rel}}$	n_{MM}	$\Delta\epsilon_{\text{rel}}$	$\Delta\rho_{\text{rel}}$
-10.88eV	320	5.16×10^{-9}	1.72×10^{-10}	376	3.16×10^{-13}	2.59×10^{-14}
-5.44eV	308	4.75×10^{-9}	2.43×10^{-11}	356	3.71×10^{-13}	1.48×10^{-15}
0.00eV	305	8.08×10^{-10}	9.50×10^{-13}	357	1.76×10^{-14}	2.39×10^{-17}
5.44eV	308	1.01×10^{-8}	1.22×10^{-12}	356	3.57×10^{-13}	8.05×10^{-17}
10.88eV	320	1.30×10^{-8}	1.56×10^{-13}	376	9.56×10^{-13}	1.83×10^{-17}

TABLE III: Three dimensional Anderson model with on-site disorder discussed in Section IV C. Number of matrix matrix multiplications n_{MM} , relative energy error $\Delta\epsilon_{\text{rel}}$, and relative L^1 density error $\Delta\rho_{\text{rel}}$.