wszaStochastic Vector Techniques in Ground-State Electronic Structure

Roi Baer[1]

Daniel Neuhauser[2]

Eran Rabani[3,4]

[1]Fritz Haber Center of Molecular Dynamics and Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
Email address: roi.baer@huji.ac.il

[2]Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA
Email address: dxn@chem.ucla.edu

[3]Department of Chemistry, University of California and Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[4]The Raymond and Beverly Sackler Center of Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv 69978, Israel
Email address: eran.rabani@berekeley.edu

Abstract. We review a suite of stochastic vector computational approaches for studying the electronic structure of extended condensed matter systems. These techniques help reduce algorithmic complexity, facilitate efficient parallelization, simplify computational tasks, accelerate calculations, and diminish memory requirements. While their scope is vast, we limit our study to ground-state and finite temperature density functional theory (DFT) and second-order perturbation theory. More advanced topics, such as quasiparticle (charge) and optical (neutral) excitations and higher-order processes, will be covered elsewhere. We start by explaining how to use stochastic vectors in computations, characterizing the associated statistical errors. Next, we show how to estimate the electron density in DFT and discuss highly effective techniques to reduce statistical errors. Finally, we review the use of stochastic vector techniques for calculating correlation energies within second-order Møller-Plesset perturbation theory (MP2) and its finite temperature variational form (GF2). Several appendices offer a succinct explanation of background theory and techniques. Example calculation results are presented and used to demonstrate the efficacy of the methods.

Contents

CHAPTER 1

Introduction

Stochastic orbital or vector techniques are emerging as valuable tools for studying the electronic structure of extended materials [1–3]. They provide a framework for reducing algorithmic complexity, facilitating efficient parallelization, breaking down computational tasks into smaller, more manageable parts, accelerating calculations, and diminishing memory requirements. While stochastic vector approaches share features with more traditional Quantum Monte Carlo techniques [4–14], they differ sharply by targeting much larger systems and relying on conventional density functional and many-body perturbation theories.

The primary tool central to stochastic vector techniques is the stochastic trace formula [15], which was used in the '90s to compute the density of states, absorption spectra, and dielectric constants of extended materials under tight-binding or semiempirical Hamiltonians [16–19]. In recent years, the stochastic vector approaches' scope and reach in electronic structure have increased dramatically. The cornerstone was ground-state density functional theory [1] using real-space-grids [20–23] and Gaussian basis sets [24, 25]. These techniques were later extended to include range-separated hybrid functionals [26, 27] and treat finite temperature effects for the warm dense matter regime [28–30]. In addition, the stochastic estimates of the forces on nuclei were used to drive Langevin molecular dynamics simulations to determine the structural properties of extended systems [31, 32]. The stochastic vector techniques were next used to describe ground state correlations beyond DFT [33–39], where they enabled lowering the algorithmic scaling and the memory requirements. Finally, their use in describing quasiparticle (charge) excitations [2, 40–48], optical (neutral) excitations [3, 49, 50], and other higher-order processes [51–53], has been rather remarkable.

This review presents key concepts and primary techniques of stochastic vector approaches in the electronic structure of extended systems. For clarity, we focus on ground-state calculations, either within DFT or in second-order perturbation theories; we will cover more advanced topics elsewhere. We start in Section 2 with a brief overview of probability and statistics theory, defining random variables and discussing sampling techniques for estimating their distribution parameters. Next, we define stochastic vectors and describe how they convert algebraic calculations into statistical estimations of the expected value of certain random variables. We then characterize the statistical errors associated with these procedures. Section 3 shows how to estimate, using stochastic vector approaches, the electron density directly from a given Kohn-Sham (KS) Hamiltonian. We do this in a real-space grid representation and also within atom-centered non-orthogonal basis sets. We then discuss the statistical errors and present highly effective techniques for reducing them, central to achieving chemical accuracy. Section 4, we review the use of stochastic vector techniques for calculating correlation energies within second-order

Møller-Plesset perturbation theory (MP2) and its finite temperature variational form, the second-order Born approximation to the single-particle Green's function (GF2). Finally, in Section 5, we provide a summary and a brief discussion of the conclusions.

CHAPTER 2

Basic stochastic vector techniques

The beating heart of a stochastic calculation is an algorithm for repeatedly generating specially crafted random variable samples. The variable's distribution is usually unknown, yet, its expected value measures exactly the physical quantity we are after (density, quasiparticle energy, force on a nucleus, Young modulus, etc.). The samples must be statistically independent and can be produced in parallel within a high-performance computation. The statistical analysis of the results enables us to find the intervals likely to include our target physical quantity with a known degree of confidence. This chapter reviews the basic definitions, concepts, and results that enable stochastic vector calculations in the electronic structure of extensive electronic systems.

2.1. Random variables

A probability distribution of a (real or complex) random variable r assigns to each of its possible values a probability $p(r)$ such that $\sum_r p(r) = 1$ (for continuous variables we replace the sum by an integral and $p(r)$ becomes the probability density). The *expected value* of the random variable is $\mathrm{E}[r] = \sum_r rp(r)$. Any function $f(r)$ is also a random variable, and its expected value is $\mathrm{E}[f(r)] = \sum_r f(r)\,p(r)$. The variance of the random variable is $\mathrm{Var}[r] = \mathrm{E}\left[|r - \mathrm{E}[r]|^2\right]$ and can be calculated from the equivalent expression $\mathrm{Var}[r] = \mathrm{E}\left[|r|^2\right] - |\mathrm{E}[r]|^2$. The standard deviation, $\sigma[r] = \sqrt{\mathrm{Var}[r]}$, is *the scale* by which we measure the deviation of r from its expected value $\mathrm{E}[r]$. Irrespective of the distribution $p(r)$, the probability for r to deviate from $\mathrm{E}[r]$ by more than $n\sigma$ is smaller than $\frac{1}{n^2}$ (the Chebyshev inequality in probability theory). This implies that the typical values of r tend to be in the $\sigma-$vicinity of $\mathrm{E}[r]$.

2.2. Estimating $\mathbf{E}[r]$

If we know the probability distribution $p(r)$, we can predict what values the random variable can take. But often, the situation is reversed: we do not know this distribution explicitly, but have a large set of independent random samples of the r. Then our task is to deduce the essential characteristics (such as expected value and standard deviation) of the distribution *from these samples.*

From the sampled values $r_1, r_2, \ldots, r_I$ of the random variable r we calculate the *sample mean*

$$m_I = \frac{1}{I}\sum_{\alpha=1}^{I} r_\alpha. \tag{2.2.1}$$

m_I itself is a random variable and its expected value provides an unbiased estimator for $\mathrm{E}[r]$:

$$\mathrm{E}[m_I] = \mathrm{E}[r] \tag{2.2.2}$$

with variance that decreases as the number of samples grows:

$$\mathrm{Var}[m_I] = \frac{1}{I}\mathrm{Var}[r] \Longleftrightarrow \sigma[m_I] = \frac{\sigma[r]}{\sqrt{I}} \tag{2.2.3}$$

As an example, we say that for $I = 100$ samples the scale of deviance of m_I from $\mathrm{E}[r]$ is 10 times smaller than that of r.

The formal way to estimate the uncertainty in m_I is through the concept of a level-p confidence interval. This is a range of values which includes $\mathrm{E}[r]$ with probability p. For building a confidence interval, we need to estimate $\mathrm{Var}[r]$ from the sample data. An unbiased estimator for this is the square of the *corrected sample standard deviation*

$$\mathrm{s}_I^2 = \frac{1}{I-1}\sum_{\alpha=1}^{I}(r_\alpha - m_I)^2 . \tag{2.2.4}$$

For large values of I (e.g., for $I > 30$), $\left[m_I - c\frac{s_I}{\sqrt{I}}, m_I + c\frac{s_I}{\sqrt{I}}\right]$ is a p-level confidence interval for $\mathrm{E}[r]$ where $c = 1,\ 2,\ 3$ corresponds to $p = 0.68,\ 0.95,\ 0.995$ respectively.[1]

2.3. Biased estimators: nonlinear functions of $\mathrm{E}[r]$

Often we need to assess the value of $f(\mathrm{E}[r])$, where f is a smooth nonlinear function. Since m_I is an estimator for $\mathrm{E}[r]$, then $f(m_I)$ can be used as an estimator for $f(\mathrm{E}[r])$. However, as explained below, the expected value of $f(m_I)$ is not guaranteed to equal $f(\mathrm{E}[r])$: we say that it is a *biased estimator* with a bias defined as:

$$b_{f,I} = \mathrm{E}[f(m_I)] - f(\mathrm{E}[r]) . \tag{2.3.1}$$

When I is large m_I is close to $\mathrm{E}[r]$ and we use Taylor's expansion of the function f around the point $\mathrm{E}[r]$ to learn of the bias's characteristics and sources:

$$f(m_I) \approx f(\mathrm{E}[r]) + f'(\mathrm{E}[r])(m_I - \mathrm{E}[r]) + \frac{1}{2}f''(\mathrm{E}[r])(m_I - \mathrm{E}[r])^2 + \ldots$$

Then, taking the expected value and using Eqs. (2.2.2)-(2.2.3) we find:

$$b_{f,I} \approx \frac{1}{2}f''(\mathrm{E}[r])\frac{\mathrm{Var}[r]}{I} + \ldots \tag{2.3.2}$$

We see that a combination of two factors produces the bias: (1) Fluctuations in r depending on the magnitude of its variance; and (2) the non-linearity of f at the point $\mathrm{E}[r]$, expressed as the magnitude of its second derivative. Eq. (2.3.2) shows that the bias diminishes asymptotically as I^{-1}.

[1] The theory applies to random variables m_I which are normally distributed. According to the central limit theorem, m_I indeed converges to a normally distributed variable when $I \to \infty$.

2.4. Stochastic vectors

2.4.1. The stochastic resolution of the identity and the stochastic trace formula. Stochastic column vectors, $\chi = \begin{pmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_K \end{pmatrix}$ have K independent random variables χ_k, each taking values from the set of unit modulus numbers, $|\chi_k| = 1$, with uniform probability. If the variables are real, χ_k is drawn from a uniform distribution of $\{\pm 1\}$, and if complex, χ_k is a number drawn from the complex unit circle, $\exp(i\theta)$, where θ is uniform random angle between $\{0, 2\pi\}$. Since there is a negative value $-\chi_k$ for each value of χ_k, the expected value must be zero $\mathrm{E}[\chi_k] = 0$, or as a vector identity: $\mathrm{E}[\chi] = 0$. Products like $\chi_k \chi_{k'}^*$, when $k = k'$ are equal to 1 and when $k \neq k'$ their expected value equals zero. Hence, $\mathrm{E}[\chi_k \chi_{k'}^*] = \delta_{kk'}$ is the identity matrix:

$$\mathrm{E}\left[\chi\chi^\dagger\right] = \mathbb{I}. \tag{2.4.1}$$

where $\mathbb{I}$ is the $K \times K$ unit matrix and $\chi^\dagger = (\chi_1^*, \ldots, \chi_K^*)$ is the row vector. Eq. (2.4.1), called *the stochastic resolution of the identity,* can be used to "convert" algebraic expressions into expected values of random variables. Once we have such a conversion, we use the statistical parameter estimation techniques of section (2.2) to obtain confidence intervals for the expected values.

2.4.2. The stochastic trace formula. The *stochastic trace formula* [15] concerning a $K \times K$ matrix A, expresses the trace of A as the expected value of the random variable $r = \chi^\dagger A \chi$[2]:

$$\mathrm{Tr}[A] = \mathrm{E}[r]. \tag{2.4.2}$$

A calculation of the variance of r shows that it is given by the sum of the square modulus of the non-diagonal elements of A:[3]

$$\mathrm{Var}[r] = \begin{cases} \sum_{k<k'} |A_{kk'} + A_{k'k}|^2 & \chi \in \mathbb{R}^K \\ \sum_{k \neq k'} |A_{kk'}|^2 & \chi \in \mathbb{C}^K \end{cases}. \tag{2.4.3}$$

With the above relations for the stochastic trace formula and the corresponding variance, we conclude that:

(1) The variance is zero, i.e.the stochastic trace is exact when:
   (a) A is diagonal
   (b) $\chi \in \mathbb{R}^K$ and $A_{kk'} = -A_{k'k}$ (for all $k \neq k'$).

---

[2] Proof of this claim based on Eq. (2.4.1): $\mathrm{Tr}[A] = \mathrm{Tr}[\mathbb{I}A] = \mathrm{Tr}\left[\mathrm{E}\left[\chi\chi^\dagger\right] A\right] = \mathrm{Tr}\left[\mathrm{E}\left[\chi\chi^\dagger A\right]\right] = \mathrm{E}\left[\mathrm{Tr}\left[\chi\chi^\dagger A\right]\right] = \mathrm{E}\left[\mathrm{Tr}\left[\chi^\dagger A\chi\right]\right] = \mathrm{E}\left[\chi^\dagger A\chi\right]$.

[3] The variance is $\mathrm{Var}[r] = \mathrm{E}\left[|r|^2\right] - |\mathrm{E}[r]|^2$ and $|\mathrm{E}[r]|^2 = |\mathrm{Tr}[A]|^2$. For $\chi \in \mathbb{C}^K$: $\mathrm{E}\left[|r|^2\right] = \mathrm{E}\left[(\chi^\dagger A\chi)^* (\chi^\dagger A\chi)\right] = \sum_{kk'qq'} A_{kk'}^* A_{qq'} \mathrm{E}\left[\chi_k^* \chi_{k'} \chi_q^* \chi_{q'}\right]$ and for $\chi \in \mathbb{R}^K$: $\mathrm{E}\left[|r|^2\right] = \mathrm{E}\left[(\chi^T A\chi)^* (\chi^T A\chi)\right] = \sum_{kk'qq'} A_{kk'}^* A_{qq'} \mathrm{E}\left[\chi_k \chi_{k'} \chi_q \chi_{q'}\right]$. In the former case use $\mathrm{E}\left[\chi_k^* \chi_{k'} \chi_q^* \chi_{q'}\right] = \delta_{kk'}\delta_{qq'} + \delta_{kq}\delta_{k'q'} - \delta_{kk'}\delta_{kq}\delta_{kq'}$ and in the latter $\mathrm{E}\left[\chi_k \chi_{k'} \chi_q \chi_{q'}\right] = \delta_{kk'}\delta_{qq'} + \delta_{kq}\delta_{k'q'} + \delta_{kq'}\delta_{k'q} - 2\delta_{kk'}\delta_{kq}\delta_{kq'}$. The result is Eq. (2.4.3).

(2) For real symmetric matrices the variance offered by the $\chi \in \mathbb{C}^K$ is a factor two smaller than that by $\chi \in \mathbb{R}^K$. Hence it may be beneficial to use complex $\chi's$ for estimating the trace.

(3) For Hermitean matrices and $\chi \in \mathbb{R}^K$ the variance is not affected by the imaginary part of the elements of A. Hence, it may be beneficial to use real stochastic vectors for estimating the trace of Hermitean matrices with very large imaginary parts.

2.4.3. Stochastic resolution in the presence of a metric. We sometimes need to work in a vector space with a positive-definite metric $\mathbb{S}^{-1}$, which is used to define the inner product between two vectors as $u^\dagger \mathbb{S}^{-1} v$, see Appendix A for an explanation and discussion in the context of computational quantum chemistry/physics. A *stochastic resolution of the metric* can be obtained by drawing the stochastic vectors χ from the following normal distribution

$$p_{\mathbb{S}}(\chi) = \sqrt{\det\left(\frac{\mathbb{S}}{2\pi}\right)} e^{-\frac{1}{2}\chi^T \mathbb{S} \chi}. \tag{2.4.4}$$

When $\mathbb{S}$ is sparse this can be done efficiently using the Metropolis algorithm [54]. It is possible to show that for such vectors [4],

$$\mathrm{E}\left[\chi\chi^\dagger\right] = \mathbb{S}^{-1}. \tag{2.4.5}$$

When $\mathbb{S}$ is sparse the Metropolis sampling procedure is fast . Often, we need to evaluate expressions such as $\mathrm{Tr}\left[\mathbb{S}^{-1}A\right]$. In such cases we define again $r = \chi^\dagger A \chi$, (but now χ are sampled from the distribution of Eq. (2.4.4)):

$$\mathrm{Tr}\left[\mathbb{S}^{-1}A\right] = \mathrm{E}\left[r\right]. \tag{2.4.6}$$

2.4.4. Variance reduction in the stochastic trace formula. Here, we discuss how $\mathrm{Var}\left[\chi^\dagger A \chi\right]$ of Eq. (2.4.3) can be reduced by projection. This approach allows to zoom into a specified portion of a system with reduced bias. The "zooming in" can be formulated algebraically as introducing a projection matrix $\mathcal{P}$ (e.g. it is Hermitean and $\mathcal{P}^2 = \mathcal{P}$) for the component of interest and $\mathcal{Q} = \mathbb{I} - \mathcal{P}$, the complementary projection, for the rest of the system. One can easily check that $\mathcal{Q}$ is also a projection $\mathcal{Q}^2 = \mathcal{Q}$, and that the two spaces are "orthogonal", i.e. $\mathcal{P}\mathcal{Q} = \mathcal{Q}\mathcal{P} = 0$. We use the relation $\mathrm{Tr}\left[A\right] = \mathrm{Tr}\left[\mathcal{P}A\mathcal{P}\right] + \mathrm{Tr}\left[\mathcal{Q}A\mathcal{Q}\right]$,[5] and $\mathcal{Q}$ contributions:

$$\mathrm{Tr}\left[A\right] = \mathrm{E}\left[\chi_{\mathcal{P}}^\dagger A \chi_{\mathcal{P}}\right] + \mathrm{E}\left[\chi_{Q}^\dagger A \chi_{Q}\right]. \tag{2.4.7}$$

---

[4] Note that for $\sqrt{\det\left(\frac{\mathbb{I}}{2\pi}\right)} \int e^{-\frac{1}{2}\chi^T \mathbb{I} \chi} \chi\chi^T d\chi = \mathbb{I}$. Then, in the integral $\sqrt{\det\left(\frac{\mathbb{S}}{2\pi}\right)} \int e^{-\frac{1}{2}\chi^T \mathbb{S} \chi} \chi\chi^T d\chi$, make a substitution for the integration variable $\tilde{\chi} = \mathbb{S}^{1/2}\chi$.This will lead to the integral in $\mathrm{E}\left[\chi\chi^\dagger\right] = \mathbb{S}^{-1}$.

[5] The trace of $\mathcal{Q}A\mathcal{P}$ and $\mathcal{Q}A\mathcal{P}$ is zero. For example, $\mathrm{Tr}\left[\mathcal{P}A\mathcal{Q}\right] = \mathrm{Tr}\left[A\mathcal{Q}\mathcal{P}\right] = 0$ where the first equality is due to the invariance of the trace to a cyclic permutation of the products and the second is due to the orthogonality of the projected spaces: $\mathcal{Q}\mathcal{P} = 0$.

Here, $\chi_{\mathcal{P}}$ ($\chi_{\mathcal{Q}}$) is a random vector in the $\mathcal{P}$ ($\mathcal{Q}$) space.[6] The variance associated with this procedure may be significantly reduced, as seen in example (2.1).

EXAMPLE 2.1. Here we show an example of reducing the variance using projection. Consider the stochastic trace evaluation of $A = \begin{pmatrix} 1 & -1 & 3 \\ 3 & 2 & 9 \\ 5 & 7 & 3 \end{pmatrix}$, the result is 10 and according to Eq. (2.4.3) (with real χ's) the variance is 324. Setting $\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, and evaluating the trace with Eq. (2.4.7), where $\chi_{\mathcal{P}} = \begin{pmatrix} \chi_1 \\ \chi_2 \\ 0 \end{pmatrix}$ and $\chi_{\mathcal{Q}} = \begin{pmatrix} 0 \\ 0 \\ \chi_3 \end{pmatrix}$, is equivalent to using the stochastic trace formula of (2.4.2) on the matrix $\mathcal{P}A\mathcal{P} + \mathcal{Q}A\mathcal{Q} = \begin{pmatrix} 1 & -1 & 0 \\ 3 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ for which the variance is 4. The procedure lead to a factor of 81 reduction in the variance [7].

Another way to reduce variance, is to use various types of similarity transforms. For example, if a unitary transform of A, e.g. $A' = U^\dagger A U$ results in a reduction of the magnitude of the off-diagonal elements, the stochastic trace formula $\mathrm{Tr}\,[A] = \mathrm{E}\left[\chi_U^\dagger A \chi_U\right]$ with $\chi_U = U\chi$ will show a smaller variance, since $\chi_U^\dagger A \chi_U = \chi^\dagger A' \chi$.

2.4.5. How variance changes as systems grow. In many situations, the trace operation is applied to a matrix representing some physical quantity in a physical system. As the size of the physical system increases, the dimension of matrix A increases but some essential physically related characteristics of matrix A do not change appreciably. Here we consider what to expect from using the stochastic trace formula Eq. (2.4.2) for increasingly larger system sizes.

(1) **Extensive quantities**. Here we assume that as the dimension K grows (in proportion to system size) but the matrix A remains diagonally dominant, i.e., $|A_{kk}| > \sum_{k' \neq k} |A_{kk'}|$, with real diagonal elements that fall within a finite interval $0 < A_{\min} < A_{kk} < A_{\max}$.[8] For each row k of the matrix A, we define a "radius" $q_k = \sqrt{\sum_{k' \neq k} |A_{kk'}|^2}$ when $\chi's$ are complex and $q_k = \sqrt{\sum_{k'<k} |A_{kk'} + A_{k'k}|^2}$ when they are real. Note that: q_k is bounded from above by A_{kk} and $\sigma[r] = \sqrt{\sum_k q_k^2}$.[9] From this, $\sum_{k=1}^K q_k^2 < \sum_{k=1}^K A_{kk}^2 < A_{\max}^2 K$, so $\sigma[r] < A_{\max}\sqrt{K}$. Combine this result with the

[6] For our purposes, we can take χ_P and χ_Q as projections from the same stochastic vector χ, with $\chi_P = P\chi$ and $\chi_Q = Q\chi$. Then $\chi = \chi_P + \chi_Q$ and $\mathrm{E}\left[\chi_P \chi_P^\dagger\right] = P$, $\mathrm{E}\left[\chi_Q \chi_Q^\dagger\right] = Q$ and $\mathrm{E}\left[\chi_P \chi_Q^\dagger\right] = 0$.

[7] Note, however, that reducing the variance this way involves twice as much numerical work (the matrix A was applied to two stochastic vectors instead of just one). So usually, working twice as hard for a factor 9 reduction in the standard deviation is an excellent "deal."

[8] A similar treatment could be given if $A_{min} < A_{kk} < A_{max} < 0$, i.e., if A_{kk} are all negative.

[9] For $\chi \in \mathbb{C}^K$ $q_k = \sqrt{\sum_{k' \neq k} |A_{kk'}|^2} \leq \sum_{k' \neq k} |A_{kk'}| \leq |A_{kk}|$. For $\chi \in \mathbb{R}^K$: $q_k = \sqrt{\sum_{k'<k} |A_{kk'} + A_{k'k}|^2} \leq \sum_{k'<k} |A_{kk'} + A_{k'k}| \leq \sum_{k'<k} (|A_{kk'}| + |A_{k'k}|) \leq |A_{kk}|$. In both

following lower bound of the expected value: $\mathrm{E}[r] = \sum_k A_{kk} \geq A_{\min} K$, finding that :

$$\text{relative uncertainty} = \frac{\sigma[r]}{\mathrm{E}[r]} \leq \frac{A_{\max}}{A_{\min}\sqrt{K}}. \tag{2.4.8}$$

Hence, the relative uncertainty *diminishes* as the system size grows in proportion to $\propto 1/\sqrt{K}$.

(2) **Local quantities.** The projection operators of the type discussed in section (2.4.4) can be used to define physically local observable in the physical system, which leads to a matrix $\mathcal{P}A\mathcal{P}$ of a low rank. The rank is often system size independent and thus, the corresponding variance associated with its self-fluctuation does not significantly change with the system size and the same can be said of the trace. The relative error in this trace calculation is therefore, *independent of the system size.*

2.4.6. Sampling and algorithmic complexity of the stochastic trace approach. Once the trace of operator A is represented as the expected value $\mathrm{E}[r]$ of the random variable $r = \chi^\dagger A\chi$, we use statistical methods to estimate it. For this, we take a sample of I stochastic vectors $\chi^{(\alpha)}$, $\alpha = 1, \ldots, I$, with $r_\alpha = \chi^{(\alpha)\dagger} A\chi^{(\alpha)}$, and use their mean m_I of Eq. (2.2.1) and their standard deviation s_I of Eq. (2.2.3) to determine the confidence intervals for A's trace, as explained in Section 2.2.

To appreciate the benefit of using the stochastic trace approach, we consider the algorithmic complexity of a deterministic trace operation, $\mathrm{Tr}[A] = \sum_{k=1}^{K} \left(u^{(k)}\right)^\dagger Au^{(k)}$ where $u^{(k)}$ are orthonormal vectors spanning the vector space. The scaling is cubic, $O\left(K^3\right)$, since for each vector $u^{(k)}$ the computational cost of calculating $Au^{(k)}$ is quadratic for each of the K vectors in the basis. However, when A can be applied in a linear scaling effort, using sparse matrix techniques, the overall scaling of the deterministic trace calculation is reduced to quadratic $O\left(K^2\right)$.

In contrast, the stochastic trace formula uses I stochastic vectors $\chi^{(\alpha)}$ that replace the K orthonormal vectors $u^{(k)}$ in the deterministic approach. As a result, the scaling is reduced I/K times, showing that the stochastic approach is much more efficient whenever $I \ll K$. Moreover, if I is independent of the size of the system, the computational complexity is reduced to linear scaling. However, for a given level of accuracy, I could be a function of K, and in such cases, the scaling can become either superlinear or sublinear with system size. It all depends on how the relative standard deviation $\sigma[r]/\mathrm{E}[r]$ in $r = \chi^\dagger A\chi$ changes with system size. In Section 2.4.5, we examined two cases, extensive and local quantities: we saw that for extensive thermodynamical observables, the relative standard deviation decreases as $1/\sqrt{K}$. Since the standard deviation of m_I is also proportional to $1/\sqrt{I}$ (Eq. 2.2.3), we find that for constant relative error, $I \times K$ should be constant, i.e., as K grows I can be made smaller. Hence emerges regime of sublinear scaling (in stochastic DFT, this was indeed observed [1, 21, 20, 23]). On the other hand, in local quantities, the variance does not decrease as system size grows but stays constant . The stochastic approach then is of linear scaling complexity since then I is constant as the system size increases.

---

cases, the leftmost inequality is valid because it's shorter to walk along the diagonal (the hypotenuse) in a right angled triangle than along the sides. The rightmost inequality is due to the diagonal dominance of A.

CHAPTER 3

Density functional theory with stochastic vectors

3.1. Basic stochastic density functional theory

The Kohn-Sham density functional theory (KS-DFT) is used to determine the energy and real-space density of a large many-electron system in contact with a reservoir at inverse temperature β and chemical potential μ. It does so by considering the unique system of non-interacting electrons with the same density under a one-body Kohn-Sham Hamiltonian operator $\hat{h}_{KS}$. The density at finite temperatures is then given by:

$$n(\boldsymbol{r}) = 2 \times \sum_k \left| f_k^{\beta,\mu} \psi_k(\boldsymbol{r}) \right|^2, \tag{3.1.1}$$

where $f_k^{\beta,\mu} = f(\varepsilon_k; \beta, \mu) = \left(1 + e^{\beta(\varepsilon_k - \mu)}\right)^{-1/2}$, and $\left(f_k^{\beta,\mu}\right)^2$ is the Fermi-Dirac occupation of energy levels ε_k corresponding to the eigenstate $\psi_k(\boldsymbol{r})$ of $\hat{h}_{KS}$. Note,

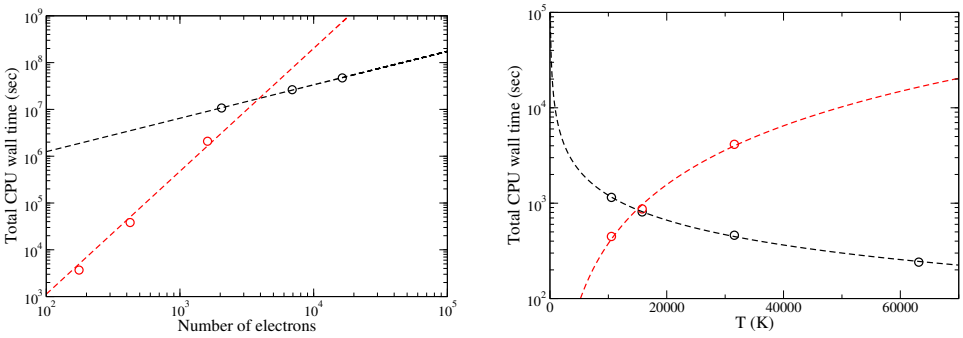

FIGURE 3.1.1. Left panel: A log-log plot of the CPU wall time as a function of the number of electrons for bulk silicon with increasing supercell size. The black symbols are results obtained from efsDFT with scaling $O\left(N_e^{0.77}\right)$ and the red symbols are deterministic DFT results with scaling $O\left(N_e^{2.7}\right)$). More details are given in Refs. [1] and [20]. Right panel: A semi-log plot of the total CPU wall time as a function of the temperature for Si_{64} using $I = 80$ stochastic orbitals (black curve) compared to deterministic calculations (red curve, using Quantum Espresso [55]) . The computational effort *increases* as $O\left(T^{2.1}\right)$ for the deterministic calculations but decreases with temperature as $O\left(T^{-0.8}\right)$ for the stochastic calculations. More details can be found in Refs. [28] and [29].

that in the limit of $\beta \to \infty$ the sum of Eq. (3.1.1) reduces to the standard expression for the ground state density, $n(\boldsymbol{r}) = 2\sum_{k=1}^{N_{occ}} |\psi_k(\boldsymbol{r})|^2$ where N_{occ} is the number of occupied KS eigenstates. The factor two in these equations is due to the dual spin occupation of the single electron eigenstate.

For numerical calculations, a finite representation is used, either a plane waves or a real-space localized basis or a real space Cartesian grid. We will concentrate here on the Cartesian grid representation, common to applications of DFT to extended materials, where the KS Hamiltonian is represented as a grid operator H.[1] The density vector at a grid point $\boldsymbol{r}_g$ is then a sum over orbital densities

$$n_g = 2 \times \sum_k \left| f_k^{\beta,\mu} \psi_g^k \right|^2 \tag{3.1.2}$$

where now $f_k^{\beta,\mu}$ are the eigenvalues (and ψ^k eigenvectors) of the square-root Fermi-Dirac operator $F^{\beta,\mu} \equiv f(H;\beta,\mu)$ and therefore can be obtained from diagonalizing the matrix H which involves $O\left(K^3\right)$ operations and $O\left(K^2\right)$ memory. The computational effort and memory requirements can be alleviated (see Fig. 3.1.1) by stochastic DFT (sDFT), involving the application of the stochastic trace formula, where the density is obtained using:

$$n_g = 2 \times \mathrm{E}\left[\left| \eta_g^{\beta,\mu} \right|^2 \right], \tag{3.1.3}$$

where

$$\eta^{\beta,\mu} = F^{\beta,\mu} \chi \tag{3.1.4}$$

and χ is a stochastic vector with elements χ_g. The rate determining step in the stochastic calculation is the operation of $F^{\beta,\mu}$ on the stochastic vector χ. This step can be executed in a linear scaling effort using a sparse representation of H while expanding the square root of the Fermi-Dirac operator $F^{\beta,\mu}$ using the Chebyshev approach, as explained in Appendix C. Moreover, unlike deterministic DFT, which requires memory storage of all occupied KS orbitals (scaling as $O\left(K^2\right)$), sDFT needs to store only three stochastic orbitals ($O(K)$) at any given time.

Besides the density, one usually needs to estimate the expected values of other one-body operators $\langle \hat{o} \rangle = 2 \times \mathrm{Tr}\left[F^{\beta,\mu} O F^{\beta,\mu} \right]$ where O is the grid representation of $\hat{o}$. Within the stochastic formulation, the quantum mechanical expectation value is converted into a statistical expected value in terms of the stochastic-projected vectors of Eq. 2

$$\langle \hat{o} \rangle = 2 \times \mathrm{E}\left[\left(\eta^{\beta,\mu} \right)^\dagger O \eta^{\beta,\mu} \right]. \tag{3.1.5}$$

Eqs. (3.1.3) and (3.1.5) also apply for other representations, for example using a basis of spatially local non-orthogonal functions $\phi_k(\boldsymbol{r})$ $(k = 1, \ldots, K)$, more common in quantum chemistry calculations. As explained in more detail in the Appendix A, a one-body operator $\hat{o}$ is represented as $O\mathbb{S}^{-1}$ in the Hilbert space spanned by ϕ_k, where $O_{kk'} = \langle \phi_k | \hat{o} | \phi_{k'} \rangle$ and the overlap matrix $\mathbb{S}_{kk'} = \langle \phi_k | \phi_{k'} \rangle$. Moreover, expectation values are given as $\langle \hat{o} \rangle = \mathrm{Tr}\left[\mathbb{S}^{-1} F^{\beta,\mu\dagger} O F^{\beta,\mu} \right]$ where $F^{\beta,\mu} \equiv$

[1] The Cartesian grid spacing is h. If $\boldsymbol{r}_g$ is a grid point then the electron density $n(\boldsymbol{r}_g)$ and KS eigenstates $\psi_k(\boldsymbol{r}_g)$ are represented by the dimensionless vectors $n_g = n(\boldsymbol{r}_g) h$ and the orthonormal vector $\psi_g^k = \psi_k(\boldsymbol{r}_g) h^{3/2}$, $\left(\sum_g \left(\psi_g^k\right)^* \psi_g^{k'} = \delta_{kk'}\right.$ and $\sum_g n_g = N_e$ is the total number of electrons in the system). The KS Hamiltonian $\hat{h}_{KS}$ is an Hermitean and linear operator on the vector space spanned by ψ^k.

$f\left(\mathbb{S}^{-1}H;\beta,\mu\right)$ (H is the matrix representing the KS Hamiltonian). The metric form of the stochastic trace in Eq. 2.4.6, Eq. (3.1.5) and Eq. (3.1.3) can be used to estimate $\langle\hat{o}\rangle$ and the density, with χ given by a stochastic column vector of K components sampled from $p_{\mathbb{S}}\left(\chi\right)=\sqrt{\det\left(\frac{\mathbb{S}}{2\pi}\right)}e^{-\frac{1}{2}\chi^{T}\mathbb{S}\chi}$ (cf., Eq. (2.4.4)).

sDFT is frequently used to study systems with a specified average number of electrons N_e. We do not have direct access to N_e since the ensemble expectation value of the number of electron operator $\left\langle\hat{N}_e\right\rangle^{\beta,\mu}=2\times\mathrm{E}\left[\left(\eta^{\beta,\mu}\right)^{\dagger}\eta^{\beta,\mu}\right]$ is a *result of the calculation*. We therefore need to find a value for μ which obeys the equation: $\left\langle\hat{N}_e\right\rangle^{\beta,\mu}=N_e$. The search for μ is done with the bisection method or the Newton-Raphson method and involves repeated evaluations of the left-hand side of the equation, each time with a different value of μ. It would be very time consuming to have to recalculate $\eta^{\beta,\mu}$ every time we change μ. In order to address this problem we use the technique of power or Chebyshev moments, developed in Refs. [56, 16, 57]. The basic idea is to expedite repeated evaluations of traces of functions of the Hamiltonian, e.g. the average electron number $\left\langle\hat{N}_e\right\rangle^{\beta,\mu}$, the KS density of states as a function of the energy $\rho\left(\varepsilon\right)=\left\langle\frac{\partial}{\partial\mu}N^{\beta\mu}\right\rangle_{\mu=\varepsilon}$ [1], and the entropy [28]. We explain how moments are used in sDFT in Appendix C.2.

3.2. Statistical errors and techniques for their reduction

The random fluctuations in the sample estimates of the electron density of Eq. (3.1.3) propagate into the density-dependent Hartree and the exchange-correlation potentials. The latter potential depends non linearly on the density and therefore exhibits a bias (see discussion in Section (2.3)). As the self-consistent electron density depends non-linearly on the fluctuating and biased KS Hamiltonian (through a Chebyshev expansion) it too is biased.

Refs. [24, 25, 28] studied the fluctuation and bias errors in sDFT demonstrating that (1) The relative fluctuations in the density reduce as systems grow in size; sources for this effect were discussed in Section (2.4.5) and (2) the relative bias in the electron density, forces and energy are small (especially when embedded fragments discussed below are used) and do not grow with system size.

Reducing statistical errors is always possible by more sampling. This however is pricey and a lot can be achieved with deterministic methods, as a preparatory steps for the sampling itself. We discuss several of these approaches below

3.2.1. Embedded fragments. An important class of techniques designed to reduce the stochastic errors is based on the embedded fragmented sDFT (efsDFT) approach [58, 31, 20, 24]. We decompose our system into F *fragments* such that each atom belongs to one fragment. The electron density in each fragment is calculated using: (1) deterministic DFT $n_g^{\mathrm{dDFT},f}$ ($f=1,\ldots,F$) and (2) a stochastic estimate $n_g^{I,f}$ of the expected value of Eq. (3.1.3), based on a sample of I stochastic vectors. We then use the sum of all fragment density differences $\Delta n_g^{I,f}\equiv n_g^{\mathrm{dDFT},f}-n_g^{I,f}$ as a correction for the stochastic estimate of the density for the entire system: $n_g^{I,\mathrm{EF}}=n_g^{I}+\sum_{f=1}^{F}\Delta n_g^{I,f}$. When $I\to\infty$, $n_g^{I,f}\to n_g^{\mathrm{dDFT},f}$ and the corrections $\Delta n_g^{I,f}$ tend to zero, while at the same time the stochastic density n_g^{I} tends to the deterministic n_g^{dDFT}. The expectation value of any operator of interest $\hat{o}$ can also

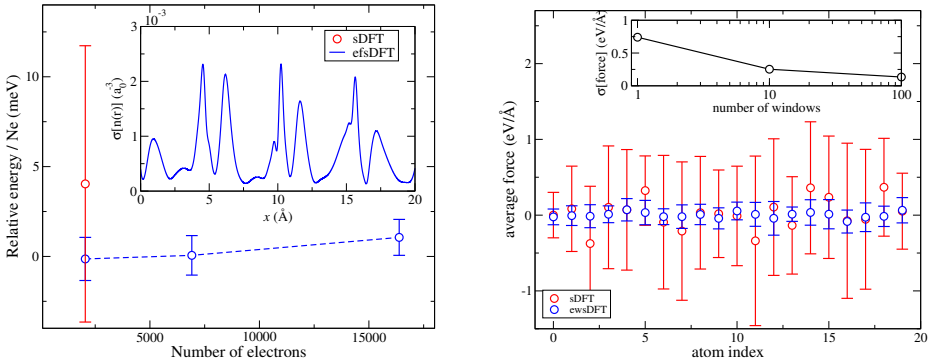

FIGURE 3.2.1. Left panel: The total energy per electron relative to a deterministic DFT calculation for bulk silicon in a supercell with $N_a = 512$ Si atoms ($N_e = 2048$ valance electrons) as a function of the number of electrons for increasing supercell size. Inset: The standard deviation of the density along the [100] direction. Red/blue curves/symbols correspond to sDFT/efsDFT calculations with $I = 800$ (red) and $I = 80$ (blue), respectively. Note the significant reduction in the standard deviation for efsDFT (much smaller error bars obtained from ten runs) for $I_{\text{efsDFT}} \ll I_{\text{sDFT}}$. More details can be found in Ref. [20]. Right panel: The expectation value of the force along the [100] direction of representative atoms for bulk silicon with $N_a = 512$ Si atoms in the supercell. The results obtained from ewsDFT with $N_w = 100$ windows (blue symbols) are compared to sDFT (red symbols) with $I = 512$ stochastic orbitals. Inset: The standard deviation of the force averaged for all atoms as a function of the number of windows N_w. The results for N_w=1 correspond to sDFT. A significant reduction in the standard deviation is obtained already for $N_w \approx 10$. More details can be found in Ref. [21].

be corrected as: $\langle\hat{o}\rangle^{I,\text{EF}} = \langle\hat{o}\rangle^{I} + \sum_f \langle\Delta\hat{o}\rangle^{I,f}$ where $\langle\Delta\hat{o}\rangle^{I,f} = \langle\hat{o}\rangle^{\text{dDFT},f} - \langle\hat{o}\rangle^{I,f}$. This scheme leads to a significant reduction in the bias and in the statistical error, as illustrated in the left panel of Fig. 3.2.1.

3.2.2. Energy windows. In the energy windows technique [21], we divide the energy axis ε into N_W "windows" by introducing a monotonic sequence of chemical potentials: $\mu_1 < \mu_2 < \cdots < \mu_{N_W} = \mu$ (the last potential is the physical chemical potential μ). With these, we define "energy window" functions

$$f^1(\varepsilon)^2 = f^{\beta\mu_1}(\varepsilon)$$
$$f^w(\varepsilon)^2 = f^{\beta\mu_w}(\varepsilon) - f^{\beta\mu_{w-1}}(\varepsilon), \quad w = 2, \ldots, N_W.$$

Note that $\sum_w f^w(\varepsilon)^2 = \left(f^{\beta\mu}\right)^2$ and thus, the KS expectation value of a one-body operator can be expressed as $\langle\hat{o}\rangle = 2 \times \text{Tr}\left[\sum_{w=1}^{N_W} F^w O F^w\right]$ where $F^w = f^w(H)$.

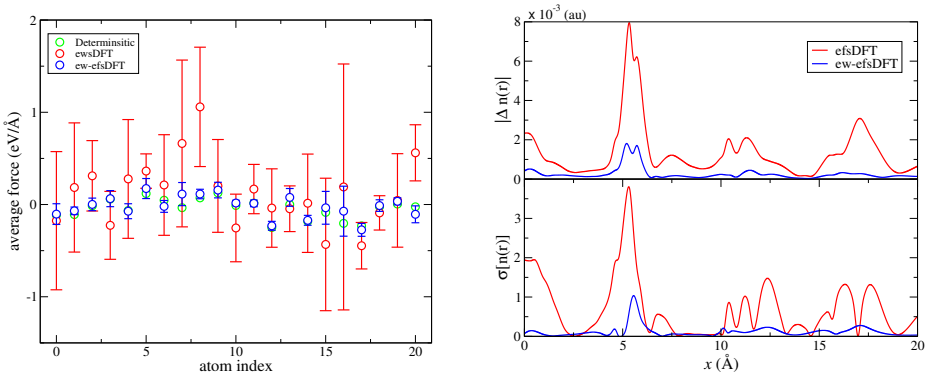

FIGURE 3.2.2. Left panel: Comparison of the expectation value of the force along the [100] direction of representative atoms for G-center impurity embedded bulk silicon with $N_a = 512$ atoms in the supercell for ewsDFT (red) and ew-efsDFT (blue). Error bars in the forces on the nuclei were obtained from five runs. Also shown are the deterministic results (green). Right panels: The absolute value of the electron density difference ($|\Delta n(r)|$) between the stochastic and a deterministic calculation (upper panel) and the standard deviation of the electron density ($\sigma[n(r)]$) (lower panel) along the [100] direction. Results for efsDFT and ew-efsDFT are shown in red and blue, respectively. More details can be found in Ref. [23]

The corresponding stochastic trace formula is then:

$$\langle \hat{o} \rangle = 2 \times \sum_{w=1}^{N_W} \mathrm{E}\left[\eta^{w^\dagger} O \eta^w\right], \tag{3.2.1}$$

where $\eta^w = F^w \chi$ and the calculation of the N_W vectors, η^w, involves a Chebyshev expansion $\eta^w = \sum_{n=0}^{N_C-1} a_n^w \chi_n$ with coefficients $\left\{a_0^w, a_1^w, \cdots, a_{N_C-1}^w\right\}$, which depend on the function $f^w(\varepsilon)$ (see C for details). We see, that each η^w has its own set of coefficients $\left\{a_0^w, a_1^w, \cdots, a_{N_C-1}^w\right\}$ but all the *η^ws share the same set of Chebyshev vectors* $\{\chi_0, \chi_1, \cdots \chi_{N_C-1}\}$. All η_w's use the same Chebyshev vectors χ_n but each requires a distinct set of coefficients a_i^w. The calculation of χ_n is by far the dominant part of the computation therefore the numerical effort is almost entirely independent of the number of windows N_w.

The above energy window scheme can be rather useful in reducing the statistical noise at a small computational cost, but only for an observable, O, that does not commute with H. This may (or may not, depending on O) yield a lower overall statistical error. When O commutes with H, the F^w's and ρ^w's commute with O and the energy windowing has absolutely no effect on the statistical error. Our previous calculations suggest that the energy window is highly useful in reducing the bias and the statistical error on the forces on nuclei [21]. An even more impressive reduction in the bias and statistical noise can be obtained when windowing is combined with the embedded fragment approach [23], as illustrated in Fig. 3.2.2.

3.2.3. Tempering. Another approach that is useful to reduce the statistical error and the bias is based on introducing a reference system which is kept at a higher temperature. This allows us to rewrite the Fermi-Dirac operator $f^{\mu\beta}(\varepsilon)$ appearing in Eq. (3.1.4) as a sum of two terms, a high temperature term ($\beta^w < \beta$) and a correction term [59]:

$$f^{\beta\mu}(\varepsilon)^2 = f^{\beta^w\mu}(\varepsilon)^2 + \Delta f(\varepsilon), \tag{3.2.2}$$

where β is the target inverse temperature. This leads to the following expression for expectation values $\langle\hat{o}\rangle^{\beta\mu} = \langle\hat{o}\rangle^{\beta^w\mu} + \Delta\langle\hat{o}\rangle$ composed of the following stochastic traces:

$$\Delta\langle\hat{o}\rangle = 2 \times \mathrm{E}\left[\eta^{\beta\mu\dagger} O \eta^{\beta\mu} - \eta^{\beta^w\mu\dagger} O \eta^{\beta^w\mu}\right]$$

and

$$\langle\hat{o}\rangle^{\beta^w\mu} = 2 \times \mathrm{E}\left[\eta^{\beta^w\mu\dagger} O \eta^{\beta^w\mu}\right].$$

In the above, $\langle\hat{o}\rangle^{\beta^w\mu}$ is the larger contributor to the expected value of $\langle\hat{o}\rangle$ and since this term $\eta^{\beta^w\mu}$ is evaluated at higher temperatures, it requires Chebyshev expansion lengths which are a factor $\approx \beta/\beta^w$ shorter than that required to evaluate $\eta^{\beta\mu}$. On the other hand, the smaller correction term ($\Delta f(\varepsilon)$) requires a small number of samples I since it overall magnitude is rather small. Thus, we use a large number of stochastic vectors to evaluate $\eta^{\beta^w\mu}$ with a short Chebyshev expansion and a small number of stochastic vectors to evaluate $\Delta\langle\hat{o}\rangle$ with a long Chebyshev expansion. This allows to reduce the overall bias and statistical error for the same computational cost.

CHAPTER 4

Stochastic vectors for weakly correlated systems beyond DFT

In this section we show how stochastic vectors can help reduce the workload or scaling in post Kohn-Sham calculations. We focus on weakly correlated systems that can be described within the second order Born approximation to the self energy within Matsubara Green's Function theory [60]. In Appendix B we present a brief review of the Matsubara Green's Function theory and its application to electronic structure theory. In brief, the central variable of the theory is the many body thermal Green's function, $G(\tau)$ which depends on a parameter $\tau \in [0, \beta]$ called "imaginary time". To compute $G(\tau)$, one requires as input the self energy, $\Sigma(\tau)$, which is the most time-intensive operation. Within the second order Born approximation the calculations of $\Sigma(\tau)$ scales as $O\left(K^5\right)$ for K basis functions, cf. Eq. (B.3.4). Here we describe how stochastic vector techniques can reduce this steep scaling to either quadratic $O\left(K^2\right)$ or cubic $O\left(K^3\right)$ scaling.

4.1. Quadratic scaling calculation of the GF2 self energy

The basic idea here is to use a grid representation for the two-electron integrals, which can be carried out in $O(N_g \log N_g)$ operations, where N_g is the number of grid-points (proportional to the system size). The method uses three independent stochastic K-vectors: ξ, η and ζ. The different steps for every τ can be summarized as follows [37]:

(1) Define $\xi' = G(\beta - \tau)\xi$, $\eta' = G(\tau)\eta$ and $\zeta' = G(\tau)\zeta$. Note that the matrix elements of the Green's function are given by the expected value: $G_{k'k}(\beta - \tau) = \mathrm{E}[\xi_k \xi'_{k'}]$. Similarly, $G_{k'k}(\beta - \tau) = \mathrm{E}[\eta_k \eta'_{k'}]$ and $G_{k'k}(\beta - \tau) = \mathrm{E}[\zeta_k \zeta'_{k'}]$. Scaling $O\left(K^2\right)$.

(2) Define wave functions on the grid, $\psi_\xi(\boldsymbol{r}) = \sum_k \xi_k \phi_k(\boldsymbol{r})$, and similarly for $\psi_{\xi'}$, ψ_η, $\psi_{\eta'}$, ψ_ζ, $\psi_{\zeta'}$. Scaling $O(N_g)$.

(3) Calculate the stochastic two-electron integral $(\gamma\eta|\xi\zeta) \equiv \int d\boldsymbol{r}_1 \phi_\gamma(\boldsymbol{r}_1)^* v_{\xi\zeta}(\boldsymbol{r_1}) \psi_\eta(\boldsymbol{r}_1)$ where $v_{\xi\zeta}(\boldsymbol{r_1}) = \int d\boldsymbol{r_2} \frac{\psi_\xi(\boldsymbol{r}_2)^* \psi_\zeta(\boldsymbol{r}_2)}{r_{12}}$. Note that the calculation of $v_{\xi\zeta}(\boldsymbol{r_1})$ involves $O(N_g \log N_g)$ complexity but the integral over $\boldsymbol{r}_1$ in $(\gamma\eta|\xi\zeta)$ is $O(1)$ for each γ (since $\phi_\gamma(\boldsymbol{r})$ is localized in space). Repeat this for $(\gamma\zeta|\xi\eta)$ and $(\gamma'\eta'|\xi'\zeta')$. Scaling $O(N_g \log N_g)$.

(4) The self energy can be represented as an expected value of direct and exchange contributions (at each τ):

$$\Sigma_{\gamma\gamma'}(\tau) = \mathrm{E}\left[-(\gamma'\eta'|\xi'\zeta')\left(2(\gamma\eta|\xi\zeta) - (\gamma\zeta|\xi\eta)\right)\right]. \tag{4.1.1}$$

Note that the only term that depends on τ is $(\gamma'\eta'|\xi'\zeta')$, this fact greatly expedites the Fourier transform of $\Sigma(\tau)$ into a frequency dependent quantity (which is required in some implementations).

Algorithmic complexity: The overall calculation of all $(\gamma\eta|\xi\zeta)$'s is of linear scaling algorithmic complexity, hence for a sample of size I, the overall effort is $I \times N_\tau \times N_g \log N_g$ (N_τ is the order to the time representations as discussed in Appendix B). The dependence of I on the system size may be an issue. For fixed relative correlation energy error, I was found independent of system size for the MP2 calculation [37]. However, for self consistent GF2, where the self energy is used to generate the GF, $G(\tau)$, a small bias was found that mildly depends on the system size, suggesting that for a fixed bias, the scaling is somewhat steeper than linear.

4.2. Stochastic vectors for the resolution of the identity (RI)

An alternative to the above approach is based on the resolution of the identity (RI), which is often used in deterministic MP2 and coupled cluster calculations [61, 62]. The resolution of identity uses atom-centered auxiliary charge distributions $\rho_r(\boldsymbol{r})$, $r = 1, \ldots, K'$ (typically K' is a factor 3 larger than K) spanning the space of basis function products $\phi_k(\boldsymbol{r})\phi_{k'}(\boldsymbol{r})$ to represent the $K^2 \times K'$ three-center $(kk'|A) \equiv \int d\boldsymbol{r}_1 d\boldsymbol{r}_2 \frac{\phi_k(\boldsymbol{r}_1)\phi_{k'}(\boldsymbol{r}_1)\rho_A(\boldsymbol{r}_2)}{r_{12}}$ and two-center $(A|B) = \int d\boldsymbol{r}_1 d\boldsymbol{r}_2 \frac{\rho_A(\boldsymbol{r}_1)\rho_B(\boldsymbol{r}_2)}{r_{12}}$ integrals, forming a positive definite symmetric $K' \times K'$ matrix, decomposed as $(A|B)^{-1} = \Lambda^T\Lambda$ where Λ is an invertible matrix. In a stochastic implementation of the resolution of identity (sRI) [36] we reduce the quintic scaling of calculating the self-energy to cubic by introducing two independent stochastic vectors ξ and $\tilde{\xi}$ of dimension K' and representing the 4-index Coulomb integral as an expected value:

(4.2.1)
$$(kk'|qq') \equiv \int d\boldsymbol{r}_1 d\boldsymbol{r}_2 \frac{\phi_k(\boldsymbol{r}_1)\phi_{k'}(\boldsymbol{r}_1)\phi_k(\boldsymbol{r}_1)\phi_{k'}(\boldsymbol{r}_1)\phi_q(\boldsymbol{r}_2)\phi_{q'}(\boldsymbol{r}_2)}{r_{12}} = E\left[R_{kk'}R_{qq'}\right]$$

with

(4.2.2)
$$R_{kk'} = \sum_A (kk'|A)\Lambda^T\xi.$$

Note, that we never calculate all two electron integrals but instead we merely plug their stochastic representation into Eq. (B.3.4) and after a few manipulations we obtain:

(4.2.3)
$$\Sigma[G](\tau) = -\mathrm{E}\left[\Gamma(\tau)\left(2\mathrm{Tr}\left[G(\beta-\tau)\Gamma(\tau)\right]\mathbb{I} - G(\beta-\tau)\Gamma(\tau)\right)\right]$$

where $\Gamma(\tau) = \frac{1}{2}\left(RG(\tau)\tilde{R}^T + \tilde{R}G(\tau)R^T\right)$ and R and $\tilde{R}$ are given by Eq. (4.2.2) with different random vectors ξ and $\tilde{\xi}$. The stochastic representation allows us to express Σ as a product of $K \times K$ matrices (operations scaling as K^3), which is the formal scaling of the stochastic self energy (see right panel of Fig. 4.2.1). Since other steps are also cubic (decomposition of $(A|B)$ matrix) there is no point in reducing the scaling further. Fig. 4.2.1 provides analysis of the correlation energy, the bias, and the statistical error in hydrogen chains (left and middle panel) and the scaling (right panel) of two flavors of stochastic GF2 (see Refs. [45] and [48] for more details).

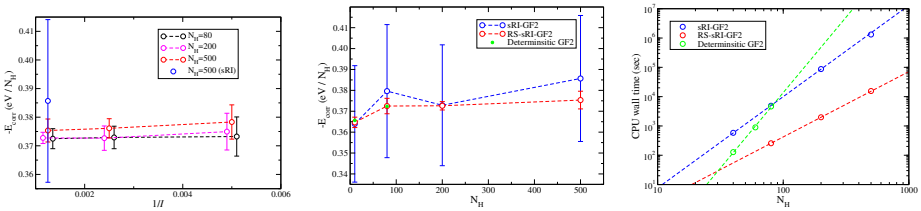

FIGURE 4.2.1. The correlation energy per electron as a function of $1/I$ for three different hydrogen chains obtained using the range-separated stochastic resolution of identity (RS-sRI) within GF2 theory. For the largest value of $I = 800$ we also show the result for the stochastic resolution of identity (sRI). Note that, for clarity, we have shifted slightly the values of the x axis for different system sizes. Middle panel: The correlation energy per electron as a function of the number of hydrogen atoms (N_{H}). The results obtained using RS-sRI (red curve) show a significant lower bias and statistical errors compared to the sRI (blue curve). Note that both approaches agree with the deterministic approach (calculated only for the smaller system sizes) within the statistical error. Right panel: Log-log plot of the computational wall time of the different GF2 approaches as a function of N_{H}. Scaling were obtained by using linear regression with $O\left(N_{\mathrm{H}}^{5.1}\right)$ for deterministic calculation, $O\left(N_{\mathrm{H}}^{2.2}\right)$ for RS-sRI-GF2, and $O\left(N_{\mathrm{H}}^{3.2}\right)$ for sRI-GF2. The standard deviation is set to 10 meV in both approaches. More details can be found in Refs. [45] and [48].

CHAPTER 5

Summary

This review presented essential concepts and techniques of stochastic vector approaches to electronic structure, focusing on the ground-state. We showed how stochastic vectors could reduce algorithmic complexity and memory requirements. For example, in sDFT, we avoided calculating the Kohn-Sham eigenstates and reduced the scaling in time and memory to linear. Furthermore, estimating the ground state correlation energy, the stochastic vector methods allowed to significantly reducing the quintic scaling of MP2/GF2 self-energy calculations.

Stochastic vector techniques enjoy many practical benefits. For example, they offer a systematic process of giving up accuracy for gaining efficiency. In addition, they provide a natural route for parallelizing calculations based on the requirement that samples must be statistically independent. However, perhaps their most significant benefit is their flexibility, allowing them to combine seamlessly with localized deterministic calculations in multiscale and embedding applications.

Acknowledgments. We thank our co-workers Eitam Arnon, Christopher Arntsen, Nadine Bradbury, Ming Chen, Yael Cytter, Wenjie Dou, Marcel Fabian, Yi Gao, Dipti Jasrasaria, Alex Lee,Wenfei Li, Minh Nguyen, John Philbin, Mykola Sereda, Ben Shpiro, Tyler Takeshita, and Vojtěch Vlček for their key contributions and hard work in developing and applying the stochastic vector methods. DN and RB are grateful to Dominika Zgid for collaborating on GF2 and to Julien Toulouse and Andreas Savin who suggested at the start of our work to apply our exchange scheme also to MP2 energies. RB, DN and ER acknowledge support from the US-Israel Binational science foundation Grant No. BSF-201836. RB acknowledges support from the Israel science foundation Grant No. ISF-800/19. DN and ER acknowledge support from the Center for Computational Study of Excited State Phenomena in Energy Materials (C2SEPEM) at the Lawrence Berkeley National Laboratory, which is funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05CH11231 as part of the Computational Materials Sciences Program. Computational resources were provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

APPENDIX A

Hilbert-space representation in a non-orthogonal basis

A Hilbert space is the complex vector space with all wave functions representing physical states of our system. Each element is a ket $|\psi\rangle$ and the space is endowed with an inner product $\langle\varphi\,|\psi\rangle \in \mathbb{C}$, for which $\langle\varphi\,|\psi\rangle = \langle\psi\,|\varphi\rangle^*$. The inner product of a function with itself is non-negative.

In computational quantum chemistry/physics we use the inner product of our system's Hilbert space to map the kets onto column vectors in a K dimensional vector space by introducing a basis of kets $|\phi_k\rangle$, $k = 1, \ldots, K$:

$$|\psi\rangle \mapsto \psi: \quad \psi_k = \langle\phi_k\,|\psi\rangle\,.$$

The inner product in the K dimensional column space is defined as following $\psi^\dagger \mathbb{S}^{-1}\varphi = \sum_{kk'} \psi_k^* \mathbb{S}^{-1}_{kk'}\varphi_{k'}$, where $\mathbb{S}^{-1}$ is called the *metric of the representation* and $\mathbb{S}_{kk'} = \langle\phi_k\,|\phi_{k'}\rangle$. The reason for this strange looking inner product is the equality of the inner product with the parent Hilbert space: if $|\psi\rangle \mapsto \psi$ and $|\varphi\rangle \mapsto \varphi$ then $\psi^\dagger \mathbb{S}^{-1}\varphi = \langle\psi\,|\phi\rangle$. We may interpret this procedure of using the metric as a transform of kets and bras:

$$\psi^\dagger \to \langle\psi|: \quad \langle\psi| = \sum_k \psi_k^\dagger \langle\phi_k|$$
$$\varphi \to |\varphi\rangle: \quad |\varphi\rangle = \sum_k |\phi_k\rangle \left(\mathbb{S}^{-1}\varphi\right)_k$$

If $\hat{o}$ is a linear operator in the Hilbert space then the matrix $O\mathbb{S}^{-1}$, where

$$O_{kk'} = \langle\phi_k\,|\hat{o}|\,\phi_{k'}\rangle \tag{A.1}$$

is a linear map in the K vector space having "the same effect" as $\hat{o}$ in the Hilbert space: if $|\psi\rangle \mapsto \psi$ and $|\varphi\rangle \mapsto \varphi$ and if $|\varphi\rangle = \hat{o}\,|\psi\rangle$ then $\varphi = O\mathbb{S}^{-1}\psi$. The quantum mechanical expectation value transforms as $\langle\psi\,|\hat{o}|\,\psi\rangle = \sum_{kk'} \psi_{k'}^\dagger \left(O\mathbb{S}^{-1}\right)_{k'k} \psi_k = \mathrm{Tr}\left[\mathbb{S}^{-1}\psi\psi^\dagger O\right]$. In finite temperature DFT the KS density matrix (KS DM) is $\hat{\rho} = \left(1 + e^{\beta(\hat{h}-\mu)}\right)^{-1}$ so in the vector space the corresponding matrix is

$$P = \left(1 + e^{\beta\left(H\mathbb{S}^{-1}-\mu\right)}\right)^{-1} \tag{A.2}$$

where, as in Eq. (A.1), $H_{kk'} = \left\langle\phi_k\left|\hat{h}_{KS}\right|\phi_{k'}\right\rangle$ and $\hat{h}_{KS}$ is the Kohn-Sham Hamiltonian. Hence

$$\langle\hat{o}\rangle = \mathrm{Tr}\left[\mathbb{S}^{-1}PO\right] \tag{A.3}$$

We can symmetrize this expression by defining $F = \left(1 + e^{\beta\left(S^{-1}H - \mu\right)}\right)^{-1/2}$ so $F^T F^T = P$ and then, using the identity $\mathbb{S}^{-1}F^T = F\mathbb{S}^{-1}$ we find

$$\langle \hat{o} \rangle = \mathrm{Tr}\left[\mathbb{S}^{-1}F^T O F\right] \tag{A.4}$$

APPENDIX B

2nd-order Matsubara Green's functions

In this section we describe the Matsubara Green's function (GF) grand canonical formalism for electronic structure and give a second order perturbation theory for its self energy.

B.1. Preliminaries

The Hamiltonian of an interacting electron system, written in second quantization

$$\hat{H} = \sum_{\sigma} \sum_{kk'} H_{kk'} \hat{a}^{\dagger}_{k\sigma} \hat{a}_{k'\sigma} + \frac{1}{2} \sum_{\sigma\sigma'} \sum_{kk'qq'} (kk'|qq') \, \hat{a}^{\dagger}_{k\sigma} \hat{a}^{\dagger}_{q\sigma'} \hat{a}_{q'\sigma'} \hat{a}_{k'\sigma}. \tag{B.1.1}$$

where $\hat{a}_{k\sigma}$ ($\hat{a}^{\dagger}_{k\sigma}$) are Fermionic annihilation (creation) operators for a σ-spin electron in the basis function $\phi_k(\boldsymbol{r})$ described in Appendix A, obeying the anti-commutation relations:

$$\hat{a}_{k\sigma}\hat{a}_{k'\sigma'} + \hat{a}_{k'\sigma'}\hat{a}_{k\sigma} = 0, \quad \hat{a}^{\dagger}_{k\sigma}\hat{a}_{k'\sigma'} + \hat{a}_{k'\sigma'}\hat{a}^{\dagger}_{k\sigma} = \delta_{\sigma\sigma'}\delta_{kk'}. \tag{B.1.2}$$

For clarity, we henceforth drop the spin designation σ where unnecessary. The matrix $H_{kk'} = \int dr \phi^*_k(\boldsymbol{r}) \left[-\frac{\hbar^2}{2m_e} \nabla^2 + u_{eN}(\boldsymbol{r}) \right] \phi_{k'}(\boldsymbol{r})$ appearing in Eq. (B.1.1), represents the one-body kinetic and electron-nuclei potential energies, while[1]

$$(kk'|qq') = \iint d\boldsymbol{r}_1 d\boldsymbol{r}_2 \frac{\phi_k(\boldsymbol{r}_1)^* \, \phi_q(\boldsymbol{r}_2)^* \, \phi_{q'}(\boldsymbol{r}_2) \, \phi_{k'}(\boldsymbol{r}_1)}{r_{12}},$$

are the *two-electron integrals* involving the pairwise electron Coulomb repulsion potential $\frac{e^2}{4\pi\epsilon_0|r_1 - \boldsymbol{r}_2|}$ which is denoted briefly as $\frac{1}{r_{12}}$. The Matsubara GF describes an open electronic system at inverse temperature β and chemical potential μ within the grand canonical ensemble. The expression $\left\langle \hat{O} \right\rangle = \mathrm{Tr}\left[\frac{e^{-\beta(\hat{H}-\mu\hat{N})}}{\mathrm{Tr}\left[e^{-\beta(\hat{H}-\mu\hat{N})}\right]} \hat{O} \right]$ defines the expectation of an operator $\hat{O}$ where $\hat{N} = \sum_{k\sigma} \hat{a}^{\dagger}_{k\sigma}\hat{a}_{k\sigma}$ is the electron number operator. For *one-body* operators $\hat{O} = \sum_{kk'\sigma} \left(\mathbb{S}^{-1}O\right)_{kk'} \hat{a}^{\dagger}_{k\sigma}\hat{a}_{k'\sigma}$ where the matrix $O_{kk'}$ is defined in Eq. A.1, and the expectation value is given in Eq. (A.3), but now

$$P^{\sigma}_{kk'} \equiv \left\langle \hat{a}^{\dagger}_{k'\sigma}\hat{a}_{k\sigma} \right\rangle \tag{B.1.3}$$

is the *interacting* density matrix (DM). For non-interacting electrons P, the KS DM, is the Fermi Dirac function of Eq. A.2.

[1] We use the notation of ref. [63], for the two-electron integrals between spatial orbitals with the following symmetries: $(kk'|qq') = (kk'|q'q)^* = (qq'|kk')$ and for real $(kk'|qq') = (kk'|q'q) = (qq'|kk') = (q'q|k'k)$.

B.2. The Fermionic Matsubara Green's function

Matsubara's GF [64] is the real symmetric matrix:

$$G_{kk'}(\tau) \equiv \theta(-\tau)\left\langle \hat{a}^{\dagger}_{k'}\hat{a}_k(\tau)\right\rangle - \theta(\tau)\left\langle \hat{a}_k(\tau)\hat{a}^{\dagger}_{k'}\right\rangle, \tag{B.2.1}$$

depending on imaginary time $-\beta < \tau < \beta$, where $\hat{X}(\tau) = e^{\tau\left(\hat{H}-\mu\hat{N}\right)}\hat{X}e^{-\tau\left(\hat{H}-\mu\hat{N}\right)}$, and $\theta(\tau)$ is 1 when τ is positive and 0 otherwise. When $\tau > 0$ the GF's matrix elements are the signed probability amplitudes for "creating an electron now and deleting it later". When $\tau < 0$ they are the amplitudes for "creating a hole in the past and deleting it now". It is readily checked that GF is anti-periodic $G(\tau) = -G(\beta+\tau)$ and thus expandable as an odd-frequency Fourier series $G(\tau) = \frac{1}{\beta}\sum_{n=-\infty}^{\infty}\tilde{G}(i\omega_n)e^{-i\omega_n\tau}$ where $\omega_n = (2n+1)\frac{\pi}{\beta}$ and $\tilde{G}(i\omega_n) = \int_0^{\beta} G(\tau)e^{i\omega_n\tau}d\tau$. For carrying out calculations, the τ-dependent functions are spanned by orthogonal polynomials up to order N_τ (typically of the order of 100-300 points). Chebyshev or Legendre expansions are possible, the latter enjoying a highly efficient convolution formula.[65, 66] From Eqs. (B.1.2), (B.2.1), and (B.1.3):

$$P = -G(\beta), \quad \mathbb{I} - P = -G\left(0^{+}\right). \tag{B.2.2}$$

One of the attractive features of the GF is the direct accessibility to the total fully interacting energy[67]:

$$E = \mathrm{Tr}\left[\dot{G}(\beta)\right] + \mathrm{Tr}\left[\mathbb{S}^{-1}HP\right] + \mu N \tag{B.2.3}$$

where here $G(\tau) = \sum_\sigma G^\sigma(\tau)$ is the "total" GF, $P = -G(\beta)$, and $N = \left\langle \hat{N}\right\rangle = \mathrm{Tr}P$.

B.3. Perturbation theory

The perturbation theory is based on a reference non-interacting Hamiltonian, with the following (so-called Fock-) matrix elements:

$$F_{kk'} = H_{kk'} + \sum_{qq'} P_{qq'}\left(2\left(kk'|qq'\right) - \left(kq'|qk'\right)\right), \tag{B.3.1}$$

for which the τ- dependent GFs take a simple "non-interacting" form:

$$g(\tau) = \left[\theta(-\tau)P - \theta(\tau)(1-P)\right]e^{-\tau\left(\mathbb{S}^{-1}F-\mu\right)} \tag{B.3.2}$$

In terms of the frequency-dependent quantities, the GF can be obtained from g by a *self-energy* matrix, $\Sigma \equiv g^{-1} - G^{-1}$ or more concretely from

$$G(\tau) = g(\tau) + \left[g \star \Sigma[G] \star G\right](\tau), \tag{B.3.3}$$

known as the Dyson equation, where the star operation is a imaginary-time convolution: $[A \star B]_{kk'}(\tau) = \sum_q \int_0^{\beta} A_{kq}(\tau-\tau')B_{qk'}(\tau')d\tau'$. The self-energy encapsulates implicitly the electron correlation, beyond the non-interacting F matrix. It can be obtained using the Luttinger-Ward variational perturbation theory [68], as a functional derivative $\Sigma(\tau) = -\frac{\delta\Phi[G]}{\delta G(\beta-\tau)}$ of a functional constructed from appropriate Feynman's diagrams. In second-order perturbation theory (summing over all pairs of identical indices) $\Phi[G] = \frac{1}{4}\int_0^{\beta} d\tau G_{p'q'}(\tau)G_{o'k'}(\tau)G_{ok}(\beta-\tau)G_{pq}(\beta-\tau) \times \left(po'|op'\right)\left(2\left(qk'|kq'\right) - \left(qq'|kk'\right)\right)$ from which:

(B.3.4)
$$\Sigma_{qp}[G](\tau) = -G_{p'q'}(\tau) G_{o'k'}(\tau) G_{ok}(\beta-\tau) (2(po'|op') - (pp'|oo')) (qk'|kq')$$

This leads to the following algorithm for obtaining the GF: start from the Hartree-Fock matrix F and the non-interacting GF g_{HF}, and set $G = g_{HF}$. Then, iterate one of the following loops until convergence:

$$\longleftarrow \begin{bmatrix} P \\ F \\ \Sigma \\ G \end{bmatrix} \xleftarrow{(B.3.1)} \begin{bmatrix} P \\ \Sigma \\ G \end{bmatrix} \xleftarrow[(B.3.4)]{(B.2.2)} G \xleftarrow[(B.3.3)]{} \begin{bmatrix} g \\ \Sigma \\ G \end{bmatrix} \xleftarrow{(B.3.2)} \begin{bmatrix} P \\ F \\ \Sigma \\ G \end{bmatrix} \longleftarrow$$

The final G and P are used to compute the energy from Eq. (B.2.3). The resulting theory is variational, respecting physical conservation laws[68]. The procedure is known as the second-order Green's function approach, GF2. If we perform just the first iteration we obtain the MP2 GF $G_{MP2}(\tau) = g_{HF}(\tau) + [g_{HF} \star \Sigma[g_{HF}] \star g_{HF}](\tau)$ and from Eq. (B.2.3) the Møller-Plesset (MP2) energy E_{MP2}, from which the MP2 correlation energy is:

(B.3.5)
$$E_{MP2}^{corr} = -\mathrm{Tr}[[\Sigma[g_{HF}] \star g_{HF}][\beta]].$$

The MP2 energy calculated for various systems is typically very close to that of the fully converged GF2 energy. Perhaps this is not a surprise in view of the variational nature of GF2.[69]

APPENDIX C

The Chebyshev expansion

C.1. Chebyshev expansion of an operator function

The Chebyshev polynomials $T_n(x)$ $(n = 0, 1, \dots)$ are a family of orthogonal polynomials in $-1 < x < 1$. In this interval the polynomials are covert *cosine* functions: $T_n(\cos\theta) = \cos(n\theta)$.

The application of a function operator, e.g $F^{\beta,\mu} \equiv f(H;\beta,\mu)$ on the stochastic vector χ carried out as a Chebyshev expansion[70, 18, 16, 71, 72]:

$$\eta^{\beta,\mu} = \sum_{n=0}^{N_C} a_n^{\beta,\mu} \chi^n \tag{C.1.1}$$

where the wave functions χ^n, result of the operation of the n^{th}'s Chebyshev polynomial $T_n(H_S)$ on the stochastic wave function χ. In a computation we compute these functions on the go, from the Chebyshev recursion

$$\chi^0 = \chi, \quad \chi^1 = H_S\chi^0, \quad \chi^{n+1} = 2H_S\chi^n - \chi^{n-1}. \tag{C.1.2}$$

The operator $H_S = \frac{H-\bar{E}}{\Delta E}$ is the shifted-scaled Hamiltonian, where ΔE and $\bar{E}$ determine an energy interval $\left[\bar{E} - \Delta E, \bar{E} + \Delta E\right]$ designed so as to contain all the eigenvalues of H. The length of the series is given by the condition $\left|a_n^{\beta,\mu}\right| < 10^{-9}$ or 10^{-10}. For efficiency, we strive to have the interval with the smallest ΔE, since $N_C \approx 4\beta\Delta E$ (depending also on μ). Finally, the expansion coefficients, depending on β and μ, are defined by $a_n^{\beta,\mu} = \frac{2}{N_C+1}\sum_{j=0}^{N_C} f\left(\Delta E\cos\theta_j + \bar{E};\beta,\mu\right)\cos(n\theta_j)$, with $\theta_j = \frac{j+\frac{1}{2}}{N_C+1}\pi$.

C.2. Chebyshev moments in sDFT

The Chebyshev moments are defined as the operator traces of the Chebyshev polynomials:[73] $M_n = \mathrm{Tr}\left[T_n(H_S)\right]$. In sDFT we use the stochastic trace formula to express each of the the moments as an expectation value of a random variable:

$$M_n = \mathrm{E}\left[\langle\chi\,|\chi_n\rangle\right], \quad n = 0, \dots, N_C. \tag{C.2.1}$$

They can be calculated in $N_C/2$ Hamiltonian applications [73] since $M_{2n} = 2\langle\chi_n\,|\chi_n\rangle - M_0$ and $M_{2n+1} = 2\langle\chi_n\,|\chi_{n+1}\rangle - M_1$ (exploiting the Chebyshev relations: $T_{n+m}(x) + T_{|n-m|}(x) = 2T_n(x)\,T_m(x)$ for $-1 \le x \le 1$).

Whenever we need to evaluate the trace of some function of the Hamiltonian $g(H,p)$ for many values of the parameter p we use Chebyshev moments to save on computational cost. For each value of p we calculate (see Section C.1.1) the Chebyshev expansion coefficients a_n^p, $n = 0, \dots, N_C$ and evaluate the trace by a

contraction with the p-independent moments:

$$\mathrm{Tr}\left[g\left(H;p\right)\right]=\mathrm{E}\left[\chi^{\dagger}g\left(H;p\right)\chi\right]=\sum_{n=0}^{N_C}a_n^pM_n. \tag{C.2.2}$$

Bibliography

[1] Roi Baer, Daniel Neuhauser, and Eran Rabani. Self-Averaging Stochastic Kohn-Sham Density-Functional Theory. *Phys. Rev. Lett.*, 111(10):106402, September 2013.

[2] Daniel Neuhauser, Yi Gao, Christopher Arntsen, Cyrus Karshenas, Eran Rabani, and Roi Baer. Breaking the Theoretical Scaling Limit for Predicting Quasiparticle Energies: The Stochastic GW Approach. *Phys. Rev. Lett.*, 113(7):076402, 2014.

[3] Eran Rabani, Roi Baer, and Daniel Neuhauser. Time-dependent stochastic Bethe-Salpeter approach. *Phys. Rev. B*, 91(23):235302, 2015.

[4] Soohaeng Yoo Willow, Kwang S. Kim, and So Hirata. Stochastic evaluation of second-order many-body perturbation energies. *The Journal of Chemical Physics*, 137(20):204122, November 2012.

[5] George H. Booth, Andreas Grüneis, Georg Kresse, and Ali Alavi. Towards an exact description of electronic wavefunctions in real solids. *Nature*, 493(7432):365–370, January 2013.

[6] James Gubernatis, Naoki Kawashima, and Philipp Werner. *Quantum Monte Carlo Methods*. Cambridge University Press, June 2016.

[7] Lucas K Wagner and David M Ceperley. Discovering correlated fermions using quantum Monte Carlo. *Rep. Prog. Phys.*, 79(9):094501, September 2016.

[8] Yann Garniron, Anthony Scemama, Pierre-François Loos, and Michel Caffarel. Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory. *J. Chem. Phys.*, 147(3):034101, July 2017. Publisher: American Institute of Physics.

[9] Guillaume Jeanmairet, Sandeep Sharma, and Ali Alavi. Stochastic multireference perturbation theory with application to the linearized coupled cluster method. *J. Chem. Phys.*, 146(4):044107, January 2017. Publisher: American Institute of Physics.

[10] Federico Becca and Sandro Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 1 edition, November 2017.

[11] J. Emiliano Deustua, Jun Shen, and Piotr Piecuch. Converging High-Level Coupled-Cluster Energetics by Monte Carlo Sampling and Moment Expansions. *Phys. Rev. Lett.*, 119(22):223003, November 2017. Publisher: American Physical Society.

[12] Alexander E. Doran and So Hirata. Monte Carlo Second- and Third-Order Many-Body Green's Function Methods with Frequency-Dependent, Nondiagonal Self-Energy. *J. Chem. Theory Comput.*, page acs.jctc.9b00693, October 2019.

[13] Maria-Andreea Filip, Charles J. C. Scott, and Alex J. W. Thom. Multireference Stochastic Coupled Cluster. *J. Chem. Theory Comput.*, 15(12):6625–6635,

December 2019. Publisher: American Chemical Society.

[14] W.A. Jr. Lester, S. Rothstein, and S. Tanaka. *Recent Advances in Quantum Monte Carlo Methods II*. World Scientific, Singapore, 2002.

[15] Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun Stat Simul Comput.*, 19(2):433–450, 1990.

[16] Lin-Wang Wang. Calculating the density of states and optical-absorption spectra of large quantum systems by the plane-wave moments method. *Phys. Rev. B*, 49(15):10154, 1994.

[17] Lin-Wang Wang and Alex Zunger. Dielectric Constants of Silicon Quantum Dots. *Phys. Rev. Lett.*, 73(7):1039–1042, August 1994.

[18] Otto F Sankey, David A Drabold, and Andrew Gibson. Projected random vectors and the recursion method in the electronic-structure problem. *Phys. Rev. B*, 50(3):1376, 1994.

[19] Toshiaki Iitaka, Shintaro Nomura, Hideki Hirayama, Xinwei Zhao, Yoshinobu Aoyagi, and Takuo Sugano. Calculating the linear response functions of noninteracting electrons with a time-dependent Schrödinger equation. *Phys. Rev. E*, 56(1):1222–1229, July 1997.

[20] Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani. Overlapped embedded fragment stochastic density functional theory for covalently-bonded materials. *J. Chem. Phys.*, 150(3):034106, January 2019.

[21] Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani. Energy window stochastic density functional theory. *J. Chem. Phys.*, 151(11):114116, September 2019.

[22] Wenfei Li, Ming Chen, Eran Rabani, Roi Baer, and Daniel Neuhauser. Stochastic embedding DFT: Theory and application to p-nitroaniline in water. *J. Chem. Phys.*, 151(17):174115, November 2019.

[23] Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani. Stochastic density functional theory: Real- and energy-space fragmentation for noise reduction. *J. Chem. Phys.*, 154(20):204108, May 2021.

[24] Marcel D. Fabian, Ben Shpiro, Eran Rabani, Daniel Neuhauser, and Roi Baer. Stochastic density functional theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10.1002/wcms.1412(0):e1412, 2019.

[25] Ben Shpiro, Marcel David Fabian, Eran Rabani, and Roi Baer. Forces from stochastic density functional theory under nonorthogonal atom-centered basis sets, 2021. _eprint: 2108.06770.

[26] Daniel Neuhauser, Eran Rabani, Yael Cytter, and Roi Baer. Stochastic Optimally Tuned Range-Separated Hybrid Density Functional Theory. *J. Phys. Chem. A*, 120(19):3071–3078, May 2016.

[27] Alex J. Lee, Ming Chen, Wenfei Li, Daniel Neuhauser, Roi Baer, and Eran Rabani. Dopant levels in large nanocrystals using stochastic optimally tuned range-separated hybrid density functional theory. *Phys. Rev. B*, 102(3):035112, July 2020.

[28] Yael Cytter, Eran Rabani, Daniel Neuhauser, and Roi Baer. Stochastic Density Functional Theory at Finite Temperatures. *Phys. Rev. B*, 97:115207, 2018.

[29] Yael Cytter, Eran Rabani, Daniel Neuhauser, Martin Preising, Ronald Redmer, and Roi Baer. Transition to metallization in warm dense helium-hydrogen mixtures using stochastic density functional theory within the Kubo-Greenwood formalism. *Physical Review B*, 100(19), November 2019.

[30] A. J. White and L. A. Collins. Fast and Universal Kohn-Sham Density Functional Theory Algorithm for Warm Dense Matter to Hot Dense Plasma. *Phys. Rev. Lett.*, 125(5):055002, July 2020.

[31] Eitam Arnon, Eran Rabani, Daniel Neuhauser, and Roi Baer. Equilibrium configurations of large nanostructures using the embedded saturated-fragments stochastic density functional theory. *The Journal of Chemical Physics*, 146(22):224111, June 2017.

[32] Eitam Arnon, Eran Rabani, Daniel Neuhauser, and Roi Baer. Efficient Langevin dynamics for "noisy" forces. *J. Chem. Phys.*, 152(16):161103, April 2020.

[33] Qinghui Ge, Yi Gao, Roi Baer, Eran Rabani, and Daniel Neuhauser. A Guided Stochastic Energy-Domain Formulation of the Second Order Møller–Plesset Perturbation Theory. *J. Phys. Chem. Lett.*, 5(1):185–189, 2013.

[34] Daniel Neuhauser, Eran Rabani, and Roi Baer. Expeditious Stochastic Approach for MP2 Energies in Large Electronic Systems. *J. Chem. Theory Comput.*, 9(1):24–27, 2013.

[35] Daniel Neuhauser, Eran Rabani, and Roi Baer. Expeditious Stochastic Calculation of Random-Phase Approximation Energies for Thousands of Electrons in Three Dimensions. *J. Phys. Chem. Lett.*, 4(7):1172–1176, 2013.

[36] Tyler Y. Takeshita, Wibe A. de Jong, Daniel Neuhauser, Roi Baer, and Eran Rabani. Stochastic Formulation of the Resolution of Identity: Application to Second Order Møller–Plesset Perturbation Theory. *J. Chem. Theory Comput.*, 13(0):4605, 2017.

[37] Daniel Neuhauser, Roi Baer, and Dominika Zgid. Stochastic self-consistent second-order Green's function method for correlation energies of large electronic systems. *J. Chem. Theory Comput.*, 13:5396–5403, 2017.

[38] Tobias Schäfer, Benjamin Ramberger, and Georg Kresse. Laplace transformed MP2 for three dimensional periodic materials using stochastic orbitals in the plane wave basis and correlated sampling. *The Journal of Chemical Physics*, 148(6):064103, February 2018.

[39] Tyler Y. Takeshita, Wenjie Dou, Daniel G. A. Smith, Wibe A. de Jong, Roi Baer, Daniel Neuhauser, and Eran Rabani. Stochastic resolution of identity second-order Matsubara Green's function theory. *J. Chem. Phys.*, 151(4):044114, July 2019. Publisher: American Institute of Physics.

[40] Vojtěch Vlček, Eran Rabani, Daniel Neuhauser, and Roi Baer. Stochastic GW Calculations for Molecules. *Journal of Chemical Theory and Computation*, 13(10):4997–5003, October 2017.

[41] Vojtěch Vlček, Wenfei Li, Roi Baer, Eran Rabani, and Daniel Neuhauser. Swift G W beyond 10,000 electrons using sparse stochastic compression. *Phys. Rev. B*, 98(7):075107, August 2018.

[42] Vojtěch Vlček, Eran Rabani, and Daniel Neuhauser. Quasiparticle spectra from molecules to bulk. *Phys. Rev. Materials*, 2(3):030801, March 2018. Publisher: American Physical Society.

[43] Vojtěch Vlček, Roi Baer, Eran Rabani, and Daniel Neuhauser. Simple eigenvalue-self-consistent δ^{-}GW0. *J. Chem. Phys.*, 149(17):174107, November 2018. Publisher: American Institute of Physics.

[44] Vojtěch Vlček. Stochastic Vertex Corrections: Linear Scaling Methods for Accurate Quasiparticle Energies. *J. Chem. Theory Comput.*, 15(11):6254–6266, November 2019. Publisher: American Chemical Society.

[45] Wenjie Dou, Tyler Y. Takeshita, Ming Chen, Roi Baer, Daniel Neuhauser, and Eran Rabani. Stochastic Resolution of Identity for Real-Time Second-Order Green's Function: Ionization Potential and Quasi-Particle Spectrum. *J. Chem. Theory Comput.*, October 2019.

[46] Vojtěch Vlček, Eran Rabani, Roi Baer, and Daniel Neuhauser. Nonmonotonic band gap evolution in bent phosphorene nanosheets. *Phys. Rev. Materials*, 3(6):064601, June 2019.

[47] Jacob Brooks, Guorong Weng, Stephanie Taylor, and Vojtech Vlcek. Stochastic many-body perturbation theory for Moiré states in twisted bilayer phosphorene. *J. Phys.: Condens. Matter*, 32(23):234001, March 2020. Publisher: IOP Publishing.

[48] Wenjie Dou, Ming Chen, Tyler Y. Takeshita, Roi Baer, Daniel Neuhauser, and Eran Rabani. Range-separated stochastic resolution of identity: Formulation and application to second-order Green's function theory. *J. Chem. Phys.*, 153(7):074113, August 2020.

[49] Yi Gao, Daniel Neuhauser, Roi Baer, and Eran Rabani. Sublinear scaling for time-dependent stochastic density functional theory. *J. Chem. Phys.*, 142(3):034106, 2015.

[50] Xu Zhang, Gang Lu, Roi Baer, Eran Rabani, and Daniel Neuhauser. Linear-Response Time-Dependent Density Functional Theory with Stochastic Range-Separated Hybrids. *J. Chem. Theory Comput.*, 16(2):1064–1072, February 2020.

[51] Roi Baer and Eran Rabani. Expeditious stochastic calculation of multiexciton generation rates in semiconductor nanocrystals. *Nano Lett.*, 12(4):2123–2128, 2012.

[52] Hagai Eshet, Roi Baer, Daniel Neuhauser, and Eran Rabani. Multiexciton Generation in Seeded Nanorods. *J. Phys. Chem. Lett.*, 5(15):2580–2585, 2014.

[53] John P. Philbin and Eran Rabani. Auger Recombination Lifetime Scaling for Type I and Quasi-Type II Core/Shell Quantum Dots. *J. Phys. Chem. Lett.*, 11(13):5132–5138, July 2020.

[54] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087, 1953.

[55] Paolo Giannozzi, Stefano Baroni, and others. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502, 2009.

[56] David A. Drabold and Otto F. Sankey. Maximum entropy approach for linear scaling in the electronic structure problem. *Phys. Rev. Lett.*, 70(23):3631–3634, June 1993.

[57] R. N. Silver and H. Röder. Calculation of densities of states and spectral functions by Chebyshev recursion and maximum entropy. *Phys. Rev. E*, 56(4):4822–4829, October 1997.

[58] Daniel Neuhauser, Roi Baer, and Eran Rabani. Communication: Embedded fragment stochastic density functional theory. *J. Chem. Phys.*, 141(4):041102, 2014.

[59] Minh Nguyen, Wenfei Li, Yangtao Li, Roi Baer, Eran Rabani, and Daniel Neuhauser. Tempering stochastic density functional theory. *arXiv:2107.06218 [physics]*, July 2021. arXiv: 2107.06218.

[60] Gianluca Stefanucci and Robert van Leeuwen. *Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction*. Cambridge University Press, 2013.

[61] Martin Feyereisen, George Fitzgerald, and Andrew Komornicki. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chemical Physics Letters*, 208(5-6):359–363, June 1993.

[62] Florian Weigend, Marco Häser, Holger Patzelt, and Reinhart Ahlrichs. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chemical Physics Letters*, 294(1-3):143–152, September 1998.

[63] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 1996.

[64] Takeo Matsubara. A New Approach to Quantum-Statistical Mechanics. *Progress of Theoretical Physics*, 14(4):351–378, October 1955.

[65] Xinyang Dong, Dominika Zgid, Emanuel Gull, and Hugo U. R. Strand. Legendre-spectral Dyson equation solver with super-exponential convergence. *J. Chem. Phys.*, 152(13):134107, April 2020. arXiv: 2001.11603.

[66] Emanuel Gull, Sergei Iskakov, Igor Krivenko, Alexander A. Rusakov, and Dominika Zgid. Chebyshev polynomial representation of imaginary time response functions. *Phys. Rev. B*, 98(7):075127, August 2018. arXiv: 1805.03521.

[67] V M Galitskii and A B Migdal. APPLICATION OF QUANTUM FIELD THEORY METHODS TO THE MANY BODY PROBLEM. *Sov. Phys. JETP*, 34(7):96–104, July 1958.

[68] J. M. Luttinger and J. C. Ward. Ground-State Energy of a Many-Fermion System. II. *Phys. Rev.*, 118(5):1417–1427, June 1960.

[69] Nils Erik Dahlen and Ulf von Barth. Variational second-order Møller–Plesset theory based on the Luttinger–Ward functional. *The Journal of Chemical Physics*, 120(15):6826–6831, April 2004.

[70] R. Kosloff. Time-Dependent Quantum-Mechanical Methods for Molecular-Dynamics. *J. Phys. Chem.*, 92(8):2087–2100, 1988.

[71] H. Röder, R. N. Silver, D. A. Drabold, and Jian Jun Dong. Kernel polynomial method for a nonorthogonal electronic-structure calculation of amorphous diamond. *Phys. Rev. B*, 55(23):15382–15385, June 1997.

[72] Roi Baer, Tamar Seideman, Shahal Ilani, and Daniel Neuhauser. *Ab initio* study of the alternating current impedance of a molecular junction. *The Journal of Chemical Physics*, 120(7):3387–3396, February 2004.

[73] RN Silver and H Röder. Calculation of densities of states and spectral functions by Chebyshev recursion and maximum entropy. *Phys. Rev. E*, 56(4):4822, 1997.