Rigorous \textit{ab initio} quantum embedding for quantum chemistry using Green’s function theory: screened interaction, non-local self-energy relaxation, orbital basis, and chemical accuracy

Tran Nguyen Lan,$^{1,2,*}$ Alexei A. Kananenka,$^1$ and Dominika Zgid$^1$

$^1$Department of Chemistry, University of Michigan, Ann Arbor, Michigan, 48109, USA
$^2$Department of Physics, University of Michigan, Ann Arbor, Michigan, 48109, USA
(Dated: June 27, 2016)

We present a detailed discussion of self-energy embedding theory (SEET) which is a quantum embedding scheme allowing us to describe a chosen subsystem very accurately while keeping the description of the environment at a lower cost. We apply SEET to molecular examples where commonly our chosen subsystem is made out of a set of strongly correlated orbitals while the weakly correlated orbitals constitute an environment. Such a self-energy separation is very general and to make this procedure applicable to multiple systems a detailed and practical procedure for the evaluation of the system and environment self-energy is necessary. We list all the intricacies for one of the possible procedures while focusing our discussion on many practical implementation aspects such as the choice of best orbital basis, impurity solver, and many steps necessary to reach chemical accuracy. Finally, on a set of carefully chosen molecular examples, we demonstrate that SEET which is a controlled, systematically improvable Green’s function method can be as accurate as established wavefunction quantum chemistry methods.

I. INTRODUCTION

In molecular systems the interplay between the localized (strongly correlated) and delocalized (weakly correlated) electrons is the chief reason causing a difficulty in describing these systems since a robust quantum chemistry method has to be able to treat both types of electrons simultaneously with a comparable effectiveness and yield molecular energies reaching chemical accuracy. Similar challenges are faced by condensed matter methods where both itinerant and localized electrons present in a solid have to be accounted for by a computational method that yields thermodynamic quantities and spectra.

In quantum chemistry, complete active space second order perturbation theory (CASPT2)$^{1-2}$, \textit{n}-electron valence state second-order perturbation theory (NEVPT2)$^{3,4}$, multi-reference coupled cluster (MRCC)$^{5-10}$, multi-reference configuration interaction (MRCI)$^{11}$, and multi-configurational density functional theory (MC-DFT)$^{12-16}$ belong to a class of methods capable of addressing the problem of strongly correlated electrons in the presence of weakly correlated ones. These are primarily wavefunction methods that are at least in principle systematically improvable and aim to produce very accurate energies for the ground state and couple of low-lying excited states.

In condensed matter community when solids are described by a realistic Hamiltonian, dynamical mean-field theory (DMFT)$^{17-19}$ on top of GW$^{20}$ method (GW+DMFT)$^{21,22}$ formulated in many-body Green’s function language is the primary tool to describe the intricate nature of realistic systems and give information about the spectral function and thermodynamic properties.

A natural direction is to find a common ground between approaches present in different communities and create a method that could yield not only thermodynamic quantities and spectra but also a satisfactory chemical accuracy while using the Green’s function language. We have developed the self-energy embedding theory (SEET)$^{23,24}$ where the self-energy describing couple of strongly correlated orbitals is embedded into the self-energy coming from the weakly correlated orbitals present in the physical problem. Consequently, SEET is a hybrid approach where we use two Green’s function methods, a weekly correlated and strongly correlated ones. The second order Green’s function (GF2)$^{25,26}$ method is used as a weakly correlated approach presenting a useful compromise between the computational cost and accuracy. We have demonstrated that GF2 can describe moderately correlated systems well$^{25}$, can be successfully used to find strongly correlated orbitals, and can deliver information about periodic systems$^{27}$ and their thermodynamic properties$^{28}$. The accurate correlated method that is used to describe a chosen set of strongly correlated orbitals is in our case a quantum impurity solver based on configuration interaction (CI) with various levels of wavefunction truncation$^{29,30}$.

We have applied SEET to the 2D Hubbard model$^{23}$ and small molecular problems$^{24}$ obtaining in both cases excellent results when compared against other established methods. For the 2D Hubbard model at half- and away from half-filling, SEET self-energies matched very well the self-energies obtained from the continuous time auxiliary field quantum Monte Carlo (CT-QMC)$^{31,32}$ calculations at all interaction strengths analyzed. In Ref. 24 using molecular examples, we have demonstrated that SEET, which is a perturb and diagonalize scheme, can reach a level of accuracy that is comparable to other quantum chemistry methods such as NEVPT2 and CASPT2. Moreover, we have shown that SEET has potential to be extended to periodic systems since in contrast to NEVPT2/CASPT2 it does not require preparing one-, two-, three- or four-body reduced density matrices and only one-body Green’s function and self-energy are sufficient to perform all calculations. Additionally, SEET in the CI-in-GF2 variant does not suffer from the intruder state problem resulting in denominator divergences present in CASPT2.

In this paper, we are further examining all the aspects of SEET and we are focusing on describing all the requirements that have to be fulfilled to make SEET systematically improvable and applicable to a wide variety of systems. We base
our discussion on the molecular systems where the near-exact (or very accurate) solutions at zero temperature are known from variety of methods such as density matrix renormalization group (DMRG)\textsuperscript{33-39} or NEVPT2/CASPT2. These simple benchmark systems allow us to establish the best way of treating molecular systems using Green’s function methods.

This paper is organized in the following way. In Sec. II A, we present a theoretical motivation behind SEET. In Sec. II B, we describe the GF2 theory that allows us to produce the self-energy for the weakly correlated orbitals. Subsequently, in Sec. II C, we proceed to explaining how SEET is related to DMFT-type approaches that use on-site effective interactions. SEET can be executed using various strongly correlated orbital selection schemes either based on energy or spatial criteria as discussed in Sec. II D. To ensure that very high accuracy and systematic improbability is maintained during SEET calculations, it is essential to use Green’s function quantum impurity solvers that can describe full realistic Hamiltonian in the impurity orbitals. We discuss possible implication of this fact in Sec. II E. Finally, we present numerical results in Sec. III and conclude this paper in Sec. IV.

II. THEORETICAL DESCRIPTION

A. Self-energy embedding theory (SEET)

Let us assume that we would like to calculate properties of a realistic system described by a Hamiltonian

\[
H = \sum_{ij} t_{ij} a_i^\dagger a_j + \frac{1}{2} \sum_{ijkl} v_{ijkl} a_i^\dagger a_j^\dagger a_k a_l,
\]

where \( t_{ij} \) and \( v_{ijkl} \) are the one- and two-body integrals in an orbital basis that can be either orthogonal or non-orthogonal. Many realistic systems are correlated enough that low-level many-body methods cannot describe them with sufficient accuracy. While high-level many-body methods can deliver accurate answers, for many interesting systems the total number of orbitals \( N \) may be too large to compute the whole problem using a high-level method. However, for most realistic cases only few orbitals contribute significantly to the physics of the problem. Consequently, these physically or chemically important orbitals, which we will call active or strongly correlated orbitals, can be described by a high-level method, while all the other inactive or weakly correlated orbitals can be described by a lower level method.

The separation of the orbital space into active and inactive or strongly and weakly correlated orbitals implies in the Green’s function language that we will express the self-energy of the strongly correlated orbitals \( u \) and \( v \) as

\[
[\Sigma]_{uv} = [\Sigma]^{low\text{-level}}_{uv} + [\Sigma]^{high\text{-level}}_{uv},
\]

where a given self-energy matrix element \([\Sigma]_{uv}\) is composed out of the local self-energy described by a high-level theory embedded into the self-energy described by a low-level theory. We assume that the self-energy contains both the frequency dependent and independent parts, \( \Sigma = \Sigma^{\omega} + \Sigma^{(i\omega)} \).

Note that the separation of the self-energy presented in Eq. 2 is characteristic for embedding methods and thus is general without necessarily specifying how the low- and high-level self-energies are evaluated in practice. Here, few major routes exist (i) either by a variational minimization of the free energy functional (that depends on the self-energy and Green’s function), (ii) by an explicit construction of the diagrammatic series necessary to represent different parts of the self-energy, (iii) or by perturbative construction that allows for updating the hybridization between the embedded system and the environment. Clearly while multitude of approximations are possible, the choice of a particular route should depend on the physics of the system under study.

In our two previous papers\textsuperscript{23,24} and in this paper, we use the third option and construct an impurity model for the strongly correlated orbitals that then has its self-energy evaluated in DMFT-type iterations; however, we would like to stress that the self-energy partitioning is general and our choice of evaluating the two parts of self-energy is one of many possible choices. The separation of the self-energy contributions allows us to evaluate the non-local part of the self-energy at a perturbative, inexpensive level, while the local, strongly correlated part is expressed as an impurity problem and is evaluated with a more accurate method.

Consequently, one of the most important questions is how to choose orbitals that are strongly correlated and ought to be treated by a higher level method. To make such a choice based on the physics and not only our intuition, we first perform a low-level perturbative calculation and analyze occupations, energy, or spatial domains of the orbitals involved. The details of such a procedure will be discussed in subsection II D.

For now, let us assume that using one of the possible criteria (orbital occupations, energies, or spatial domains), we have chosen the active (strongly correlated) orbitals that we denote \( u, v, t, w, \ldots \), while \( \mu, \lambda \) are the inactive (weakly correlated) orbitals, and \( b \) stands for a bath orbital index. We define the active space Green’s function as

\[
G_{act}(i\omega) = \left[ i\omega 1 - f_{act}^{\mu,\lambda} - \Delta(i\omega) - \Sigma_{act}(i\omega) \right]^{-1},
\]

where subscript \( \text{act} \) stands for the active space. A molecular Fock matrix without double counting is denoted as \( f^{\mu,\lambda} \). The double counting between the mean-field and high-level treatments in the active space is removed since the mean-field Coulomb interaction in the active space is exactly subtracted. Explicitly, the above operation is given by

\[
f_{uv}^{\mu,\lambda} = h_{uv} + \sum_{\mu,\lambda} P_{\mu,\lambda} \left( V_{\mu,\lambda, uv} - \frac{1}{2} V_{\mu,uv} \right) - \sum_{\lambda,\lambda} P_{\lambda,\lambda} \left( V_{\lambda,\lambda, uv} - \frac{1}{2} V_{\lambda,uv} \right),
\]

where \( P \) is the one-body density matrix obtained from a perturbative method. Note that the chemical potential has been included in the one-body electron integral matrix, \( h \), for convenience.

SEET is general and multiple perturbative methods such as GF2, GW, or FLEX\textsuperscript{40,41}, etc. can be used to describe the inactive orbitals and deliver initial density and Fock matrices.
describing all the orbitals. We have chosen to use GF2 since it contains an exchange diagram that is important for reaching the chemical accuracy in molecular systems. The details of the GF2 procedure will be listed in subsection II B. In GF2, the one-body density matrix \( P \) is obtained using Eq. 21.

The hybridization \( \Delta(\omega) \) from Eq. 3 describes the coupling of the active orbitals to the remaining weakly correlated ones. While at later iterations the active space Green’s function and self-energy are evaluated by a more accurate many-body methods, at the first iteration of SEET, to initialize these matrices, we express them as sub-matrices of the GF2 Green’s function and self-energy in the subset of active orbitals,

\[
G_{act}(\omega) = [G^{GF2}(\omega)]_{act} \tag{5}
\]

\[
\Sigma_{act}(\omega) = [\Sigma^{GF2}(\omega)]_{act}. \tag{6}
\]

The initial hybridization is then computed by substituting \( G_{act}(\omega) \) and \( \Sigma_{act}(\omega) \) into Eq. 3.

The infinite, continuum bath that describes the hybridization can be approximated by a finite, discrete one. A finite number of bath orbital energies \( \epsilon \) and the impurity-bath couplings \( V \) are then fitted to the hybridization \( \Delta(\omega) \),

\[
\Delta_{uv}(\omega) = \sum_b \frac{V_{ub}^* V_{vb}}{\omega - \epsilon_b} \tag{7}
\]

in order to produce the impurity Hamiltonian. To compute the active space Green’s function from Eq. 3 using quantum impurity solvers that need an explicit Hamiltonian formulation, one needs to formulate the Anderson impurity Hamiltonian as follows\(^{12} \)

\[
H_{act+bath} = H_{act} + H_{coupling} + H_{bath}, \tag{8}
\]

where \( H_{act} \) is the full active space Hamiltonian

\[
H_{act} = \sum_{uv} f_{uv}^{moc} a_u^\dagger a_v + \frac{1}{2} \sum_{uvwu} v_{uvwu} a_u^\dagger a_v a_w a_u, \tag{9}
\]

where \( H_{bath} \) describes the non-interacting bath, and \( H_{coupling} \) stands for the coupling between the active space and non-interacting bath,

\[
H_{bath} = \sum_b \epsilon_b a_b^\dagger a_b, \tag{10}
\]

\[
H_{coupling} = \sum_{uv} V_{ub} (a_u^\dagger a_b + a_b^\dagger a_u). \tag{11}
\]

Multiple numerical methods have been developed to solve the impurity Hamiltonian from Eq. 8. Quantum chemistry approaches that use an explicit Hamiltonian formulation and properly capture strong correlations in the active space can be extended to become Green’s function impurity solvers. A discussion concerning the exact diagonalization (ED) or full configuration interaction (FCI) solvers and a later class of solvers based on a truncated configuration interaction (CI) or restricted active space configuration interaction (RASCI) approaches can be found in Refs. 29, 30, 43–46. We employ a RASCI solver in SEET if the total number of orbitals (impurity + bath) in the impurity Hamiltonian exceeds 16 orbitals and is intractable with FCI. This quantum chemistry RASCI solver that works with Hamiltonians containing general two-body interactions is important since it allows us to achieve chemical accuracy in molecular ab initio SEET.

As an alternative to the CI solvers, the continuous time quantum Monte Carlo (CT-QMC) methods, which introduce an explicit temperature-dependence can also be used with SEET. In subsection II E, we will discuss in more details how realistic interactions impact the impurity problem and solvers.

After obtaining the active space Green’s function using the impurity Hamiltonian from Eq. 8, the active space self-energy is evaluated using the Dyson equation. Combining the non-local self-energy obtained from a perturbative method (e.g. GF2, see Eq. 17) and the active space self-energy obtained from a high level many-body method, we construct the molecular self-energy for strongly correlated orbitals as

\[
[\Sigma_{mol}(\omega)]_{uv} = [\Sigma_{non-local}^{GF2}(\omega)]_{uv} + [\Sigma_{local}^{high-level}(\omega)]_{uv}, \tag{12}
\]

where the non-local part of the self-energy coming from the perturbative method does not contain any double counting of diagrams, for details see subsection II B.

The inactive orbitals are described using only the self-energy obtained from a perturbative method and consequently

\[
[\Sigma_{mol}(\omega)]_{\mu \lambda} = [\Sigma_{GF2}(\omega)]_{\mu \lambda}. \tag{13}
\]

In SEET, the impurity Hamiltonian is solved using bare interactions since the frequency dependent field coming from all the other orbitals is described by \( \Sigma_{non-local}^{GF2}(\omega) \). However, to connect to the condensed matter and materials science community that uses screened interactions in the impurity Hamiltonian, we present a discussion of how these interactions can be evaluated in SEET in subsection II C.

The new molecular Green’s function, which contains the GF2 self-energy coming from all the inactive, weakly correlated orbitals as well as the strongly correlated self-energy from the active space orbitals, is reconstructed as

\[
G_{mol}(\omega) = \left[ i\omega \mathbf{1} - \mathbf{G}_{non} - \Sigma_{mol}(\omega) \right]^{-1}. \tag{14}
\]

This \( G_{mol}(\omega) \) is an \( N \times N \times w_{max} \) matrix and has elements for all the orbitals present in the molecular problem \( N = N_{inactive} + N_{active} \).

The hybridization \( \Delta(\omega) \) present in Eq. 3 is then updated using the new Green’s function and self-energy, and subsequent iterations are performed until convergence is achieved.

In SEET, the self-consistency procedure is made out of two loops. The inner loop has DMFT-like iterations and in this loop the matrices such as \( G_{mol}, \Sigma_{mol}(\omega) \), and \( \Delta(\omega) \) are determined self-consistently. At the convergence of the DMFT-like loop, the quantities of interest, such as the density matrix, molecular energy, and density of states, can be evaluated using the converged Green’s function and self-energy.

In SEET outer loop, we use the converged Green’s function that contains the self-energy obtained using an accurate many-body method and we pass it back to the perturbative method as a zeroth-order Green’s function. Note that in the case of GF2 only a single iteration is performed in the SEET outer
loop. (Otherwise, due to the self-consistent nature of GF2, the converged GF2 Green’s function will be the same as the one obtained previously.)

An overall scheme of SEET self-consistency is summarized as follows.

1. Generate Hartree–Fock (HF) or density functional theory (DFT) Green’s function as an initial guess.

2. Perform a self-consistent GF2 calculation for the whole molecule, see Ref. 25 for details of the GF2 iterative loop.

3. Evaluate the one-body density matrix \( \mathbf{P} \) using Eq. 21 and construct a desired orbital basis such as natural orbitals, localized orbitals, or orthogonal atomic orbitals.

4. Transform all quantities from the atomic orbital (AO) basis to the new desired orbital basis.

5. Construct the impurity Hamiltonian in which two-electron term is either a subset of bare Coulomb or screened interactions in the active space.

6. Perform DMFT-like loop:

   (a) Use an impurity solver (RASCI/FCI or CT-HYB) to obtain the active space Green’s function \( \mathbf{G}_{\text{act}}(i\omega) \) and extract the active space self-energy \( \Sigma_{\text{act}}(i\omega) \). At the first iteration the hybridization \( \Delta(i\omega) \) is initialized using GF2 quantities.

   (b) Set up the molecular self-energy according to Eqs. 12-13 if a non-local self-energy is used or Eq. 18 if screened interactions are used.

   (c) Reconstruct the molecular Green’s function via the Dyson equation and adjust the chemical potential to obtain a correct electron number for the whole molecule.

   (d) Update the hybridization \( \Delta(i\omega) \) using the new molecular Green’s function and self-energy.

   (e) Go back to step 6(a) and iterate until convergence is reached.

7. Pass the converged molecular Green’s function to a GF2 calculation and perform only a single GF2 iteration.

8. Go back to step 3 and iterate until outer loop convergence is reached.

Steps 7 and 8 are optional. In the past, we have investigated a single-shot scheme of DMFT-like iterations on top of GF2 without any further self-consistent iterations (steps 7 and 8). In practical calculations, we found that when the embedding was done in the energy domain then performing single-shot calculations was almost always sufficient. The outer loop was necessary to relax the non-local self-energy when the embedding construction was executed using spatial fragments.

B. GF2 non-local self-energy

In the first step of SEET, a low-level Green’s function method is used to obtain a non-local self-energy. In our work, we employ the \textit{ab initio} self-consistent GF2 method using either HF or DFT Green’s function as an initial guess. The GF2 self-energy in the imaginary time domain reads

\[
\Sigma_{\text{mol}}^{GF2}(\tau)_{ij} = -\sum_{klmpq} G_{\text{mol}}^{GF2}(\tau)_{ik} \left[ G_{\text{mol}}^{GF2}(\tau)_{jm} \right]_{mn} \left[ G_{\text{mol}}^{GF2}(\tau)_{pq} \right]_{ij} \times v_{ikmq}(2v_{jpm} - v_{jpn}),
\]

(15)

where \( \mathbf{G}_{\text{mol}}^{GF2}(\tau) \) is the imaginary time GF2 Green’s function and

\[
v_{ijkl} = \int d\mathbf{r}_1 d\mathbf{r}_2 \phi^*_i(r_1) \phi_j(r_1) \psi(r_1 - r_2) \phi_k^*(r_2) \phi_l(r_2),
\]

(16)

are two-electron integrals in the AO basis. The resulting \( \Sigma_{\text{mol}}^{GF2}(\tau) \) matrix is then transformed to the imaginary frequency domain \( \Sigma_{\text{mol}}^{GF2}(i\omega) \) using either a traditional Fourier transform (FT) or a Fourier transform in the basis of orthogonal polynomials\(^47\). The non-local part of the GF2 self-energy without the double counting is directly defined as a difference between the GF2 self-energy obtained from Eq. 15 while summing over all the molecular orbitals and that obtained by summing only over the local (active) part

\[
\Sigma_{\text{non-local}}^{GF2}(i\omega) = \Sigma_{\text{mol}}^{GF2}(i\omega) - \Sigma_{\text{act}}^{GF2}(i\omega),
\]

(17)

where \( \Sigma_{\text{act}}^{GF2}(i\omega) \) is evaluated using local interactions within the active space. The inclusion of this non-local self-energy into the DMFT-like iterations of SEET eliminates the need for screened, effective interactions in the active space since the non-local self-energy term contains all the non-local interactions between the active and remaining orbitals. We will elaborate on this point in subsection II.C.

Recently, a comprehensive comparison between approximate diagrammatic schemes including second-order perturbation theory (\( \Sigma^2 \)), GW, FLEX, and \( T \)–matrix approximation (TMA)\(^48\) was performed for the 2D Hubbard model\(^49\) where formally exact results are known for multiple regimes. This study showed that the GW, FLEX, and TMA methods consisting of partial summations of bubble and/or ladder diagrams can yield worse results than the second-order perturbation theory. More importantly, when combined with DMFT, \( \Sigma^2 \) non-local self-energy is more reliable than that from other methods including partial diagrammatic summations\(^49\). We have chosen GF2 as the low-level perturbative method since we expect that similar conclusions will hold for realistic systems, particularly for molecular systems, where it is widely acknowledged that an exchange diagram is important if chemical accuracy is required.

The main bottleneck of GF2 is the evaluation of the second-order self-energy from Eq. 15. However, a highly parallel scheme consisting of factorizations and multiplications can easily make the GF2 calculation possible for a couple of hundred orbitals.
The other bottleneck in the GF2 calculation is the size of the imaginary frequency and time grids necessary to express the self-energy and Matsubara Green’s function. To reach chemical accuracy for realistic systems, these grids can have up to ten or even hundred thousand of points, leading to computations that are both processor and memory demanding. Recently, we have optimized these numerical grids. First, a Fourier transform in the basis of orthogonal polynomials\(^{47}\) was used to transform the second-order self-energy from the imaginary time to the Matsubara frequency domain instead of a conventional Fourier transformation. The imaginary time grid can be therefore reduced to a couple of hundred points, while preserving a micro-Hartree accuracy in the energy evaluation. Later, the cubic spline interpolation was implemented to approximate the equidistant Matsubara frequency grid\(^{50}\). We showed that the chemical accuracy can be maintained with a very sparse subset of imaginary frequency points (only a few percent of the full imaginary frequency grid). These new grids significantly lower the required computation time and memory storage for the the self-energy and Matsubara Green’s function.

C. Screened interactions in SEET

In a molecular system, the active space self-energy obtained from an impurity described by the bare interactions \(\Sigma_{\text{bare}}\) combined with the the self-energy \(\Sigma_{\text{non-local}}(\tau, \mathbf{r})\) evaluated by summing over all weakly correlated (inactive) orbitals is equivalent to the self-energy evaluated for an impurity problem described by screened interactions \(U(\tau)\). Numerically, this amounts to requiring that the impurity self-energy evaluated using imaginary time dependent screened interactions recovers the full second-order self-energy at every \(\tau\) point

\[
[\Sigma_{\text{mol}}(\tau)]_{ii} = [\Sigma_{\text{non-local}}(\tau, v_{\text{bare}})]_{ii} + [\Sigma_{\text{act}}(\tau, v_{\text{bare}})]_{ii} = [\Sigma_{\text{act}}(\tau, U(\tau))]_{ii}.
\]  

Consequently, in multiple theories present in condensed matter and materials science the non-local field coming from all other electrons is described by \(U(\tau)\), while in SEET this non-local field is described by \(\Sigma_{\text{non-local}}(\tau, v_{\text{bare}})\). Typically, screened interactions are obtained by a downfolding procedure of the full realistic band structure or infinite lattice Hamiltonian to an effective impurity model with only few correlated orbitals\(^{51-54}\). In SEET, we avoid the explicit downfolding procedure by including the \(\Sigma_{\text{non-local}}(\tau, v_{\text{bare}})\) matrix in the DMFT-like self-consistency.

However, in order to connect to the many-body condensed matter and materials science community, we demonstrate how to evaluate on-site screened interactions in molecular problems and by numerical examples (subsection III A) show that both the procedures, SEET and DMFT with screened interactions, give similar numerical answers.

In our previous work, we proposed a procedure for finding screened interactions that reached satisfactory chemical accuracy for molecular systems\(^{55}\). In this work, we capitalize on the previous procedure, and we require that for a single strongly correlated on-site orbital we fulfill the following relationship

\[
[\Sigma_{\text{mol}}^{\text{GF2}}(\tau)]_{ii} = [\Sigma_{\text{act}}^{\text{GF2}}(\tau, U(\tau))]_{ii}
\]

\[
= \left[-G_{\text{mol}}^{\text{GF2}}(\tau)\right]_{ii}^2 \left[G_{\text{mol}}^{\text{GF2}}(-\tau)\right]_{ii} [U(\tau)]^2,
\]

for the on-site GF2 imaginary time self-energy (Eq. 15). The on-site screened interaction \(U(\tau)\) is then obtained as follows

\[
U(\tau) = -\sqrt{\left[\Sigma_{\text{mol}}^{\text{GF2}}(\tau)\right]_{ii}^2 \left[G_{\text{mol}}^{\text{GF2}}(-\tau)\right]_{ii}}.
\]

It is worth mentioning that while the non-local self-energy in Eq. 17 for multiple-orbital active spaces can be straightforwardly evaluated, the generalization of the on-site screened interaction to treat multiple orbitals, unfortunately, is non-trivial. This is due to the multiple choices of screened interactions \(U_{ij}(\tau)\) and the parametrization not always being unique\(^{55}\). Additional advantages of using the non-local self-energy as compared to using the screened interaction in SEET will be numerically demonstrated in subsection III A.

Now, let us shortly discuss the main distinction between SEET using the GF2 non-local self-energy and GW+DMFT. First, in GW+DMFT, the GW non-local self-energy is given by the product of Green’s function and the dynamically screened interaction \(W\). Thus, in GW+DMFT, when the realistic band structure or infinite lattice Hamiltonian is mapped onto an effective low energy model, it is mandatory to reconstruct the local components of both \(G\) and screened interaction \(W\)\(^{21}\). This leads to the “double embedding in both Green’s function \(G\) and dynamically screened interaction \(W\)”\(^{54}\).

In SEET, the GF2 non-local self-energy is computed using bare Coulomb interactions (Eq. 15), it is therefore unnecessary to reconstruct the local component for the interaction in the active space, which can be directly taken from the bare interaction matrix. Second, because of the dynamical nature of local screened interactions, the rigorous solution the impurity problem without neglecting the \(\tau\)-dependence of the interactions in GW+DMFT can be only obtained by CT-QMC solver. In SEET, any existing quantum impurity solver can be used to tackle the impurity problem. GW+DMFT is known to yield good results for realistic materials such as NiO, it remains to be established if SEET can yield accurate results for these systems. However, for molecular systems, SEET with the GF2 non-local self-energy is a fully \textit{ab initio}, systematically improvable, and self-consistent procedure able to give results with chemical accuracy.

D. Orbital basis

After a self-consistent GF2 calculation is performed, the next step in the SEET procedure is to construct an orbital basis in which active orbitals are chosen. Here, as we have mentioned earlier, two general schemes are possible: either to perform the selection based on the orbital energies or spatial
domains criteria, depending on system studied. As an example, in Fig. 1, we present the selection of impurities using the aforementioned criteria for the H₈ chain.

**Spatial impurity selection scheme**

![Spatial impurity selection scheme](image)

**Energy or occupation impurity selection scheme**

![Energy or occupation impurity selection scheme](image)

**FIG. 1.** Two schemes for selecting orbitals used to construct impurities for the H₈ chain. Note that in the spatial selection scheme there are two pairs of degenerate impurities (A, D) and (B, C). In the energy or occupation scheme, there is no such degeneracy and impurities are build using bonding and corresponding antibonding orbitals.

To construct the orbital basis in the energy domain, the one-body density matrix **P** is first evaluated using the converged GF2 Green’s function,

$$P = -2G_{mol}^{GF2}(\tau = \beta),$$

where $\beta = 1/(k_B T)$ is the inverse temperature. Then, this whole molecular one-body density matrix is diagonalized to obtain the natural orbitals (NOs) and occupation numbers. The active orbitals are then chosen from this set of NOs based on their occupations, as it is done in traditional CAS type methods. The strongly correlated orbitals have occupations significantly different from 0 or 2, while the weakly correlated orbitals are mostly empty or doubly occupied. In SEET performed in the NO basis, if the number of active orbitals is too large to be included in a single impurity, the orbitals can be easily split into different groups (impurities) belonging to different fragments or different symmetries without any further implementation as shown in Fig. 1, for numerical examples see Table IV.

In the spatial domain, the localized orbitals can be formally obtained by localizing NOs using the well-known Pipek–Mezey and Boys localization schemes. Note that Boys orbitals are a molecular analogue of Wannier orbitals. In this work, we use the Löwdin orthogonalized AO (SAO) basis when dealing with a minimal basis set, while the so-called regional natural orbital (RNO) basis is employed for larger basis sets than the minimal one. The construction of RNO basis for SEET can be briefly described as follows. Starting from AO density matrix $P^{SAO}$, the density matrix in SAO basis $P^{SAO}$ can be obtained by the Löwdin orthogonalization,

$$P^{SAO} = S^{1/2}P^{SAO}S^{1/2},$$

where $S$ is an AO overlap matrix. In the next step of RNO basis construction, we separately diagonalize a block density matrix of predefined fragments $i$

$$P^{RNO} = U_i^{\dagger}P^{SAO}U_i.$$  \hspace{1cm} (23)

The transformation matrix from SAO to RNO basis for a molecule consisting of $n$ molecular fragments is a direct summation of all block eigenvectors $U_i$

$$U = U_1 \oplus U_2 \oplus ... \oplus U_n.$$  \hspace{1cm} (24)

The density matrix in RNO basis can be obtained as follows

$$P^{RNO} = U^{\dagger}P^{SAO}U.$$  \hspace{1cm} (25)

Finally, the RNO coefficients present in the non-orthogonal AO basis are obtained by a back transformation

$$C^{RNO} = S^{-1/2}U.$$  \hspace{1cm} (26)

The active orbitals are then chosen from RNOs belonging to particular molecular fragments.

It is worth mentioning that in the original description of the RNO construction, the RNO density matrix (Eq. 25) is further diagonalized using the Jacobi rotation to obtain the bonding between fragments. However, since our purpose is to approximately disentangle the bonding between fragments, we will not proceed according to the original description.

Generally, depending on a system under study, we can use either NO or SAO/RNO bases. If the entanglement between molecular fragments is large, the NO basis should be used to correctly describe the bonding between fragments. On the other hand, if molecular fragments are only weakly coupled, we approximately can separate them and employ SAO/RNO bases. The advantages and disadvantages of each orbital basis will be carefully demonstrated using numerical results in subsection III.C.

E. Implications for impurity solvers to reach chemical accuracy

For SEET procedure to be computationally well behaved and accurate one has to fulfill multiple requirements: (i) orthogonality of the orbital basis to properly carry out the embedding procedure when an explicit bath representation for a CI solver is necessary, (ii) including a realistic Hamiltonian in the impurity problem, (iii) hybridizations that simplify the bath fitting procedure, (iv) locality to make the self-energy decay fast with respect to the distance, and (v) a possibility of treating many strongly correlated orbitals by an impurity solver. All of these requirements have implications for the possible quantum impurity solvers that can be used in SEET or DMFT procedures.

For molecules, if chemical accuracy is desired, any modification of the realistic Hamiltonian containing full one-
two-body interactions to a simplified Hamiltonian including only a subset of all interactions may have a detrimental effect\textsuperscript{61}. Many of the off-diagonal elements of the realistic Hamiltonian can have a similar magnitude to the diagonal elements, thus presenting no justification for neglecting them and keeping only the diagonal elements. Moreover, solutions obtained by such modifications usually cannot be improved in a systematic manner. Consequently, for molecular calculations the safest option for preserving chemical accuracy is to use a class of quantum impurity solvers that employ a full realistic Hamiltonian for the impurity orbitals. This requirement allows us to use two classes of solvers: either based on CI expansions or continuous time hybridization expansion quantum Monte Carlo (CT-HYB QMC\textsuperscript{62–65}).

Solvers based on CI expansions such as restricted active space CI (RASCI) allow us to comfortably include around 8 orbitals in the impurity and around 16-24 bath orbitals. They usually work in the zero (or low) temperature formulation and require a bath discretization that can be a potential source of errors. The hybridization expansion solvers can handle an infinite bath but face difficulties with Monte Carlo sign problem that gets especially pronounced for non-diagonal hybridizations and low temperatures. Consequently, for both classes of solvers non-diagonal hybridizations lead to problems. For CI solvers non-diagonal hybridizations lead to large number of bath orbitals necessary to fit the discrete bath accurately. For hybridization expansion solvers non-diagonal hybridizations lead to long computation times due to sign problem. Thus, a crucial challenge is to define a set of orbitals that either diagonalizes hybridization or minimizes its off-diagonal elements for a range of frequencies.

To gain insight into the possibility of making the hybridization as diagonal as possible, we plot in Figs. 2 and 3 hybridizations for different impurities in the H\textsubscript{8} chain in the 6-31G basis\textsuperscript{66} using different selection schemes presented in Fig. 1. Spatially localized orbitals are obtained using the Boys localization method implemented in GAUSSIAN 09\textsuperscript{67}. It is evident that NOs yield a diagonal hybridization while Boys orbitals yield hybridization containing off-diagonal elements of large magnitudes. This insight explains some of our observations that NOs are quite advantageous for SEET calculations since the bath fitting procedure can be done using relatively few bath sites.

We also consider how differentorbitals, by influencing hybridization behavior, can change the performance of CT-HYB QMC, in Tables I and II. Let us note in passing that for a minimal basis (STO-3G\textsuperscript{68}), SAOs are very similar to Boys orbitals and the off-diagonal elements of hybridization are significant for SAOs. It is evident that CT-HYB QMC has a much worse average sign and larger perturbation order in orbital bases that yield non-diagonal hybridizations, thus making such calculations particularly challenging especially at low temperatures since the average perturbation order in CT-HYB QMC calculations grows as temperature decreases. We observe that even for high temperatures the difference in CT-HYB QMC performance is present and NOs are still favored. Moreover, this difference in performance is independent of the number of orbitals present in a given basis and the trend is maintained when a larger orbital basis such as 6-31G is employed. Our observations indicate that for molecular examples NOs are advantageous when used with the CT-HYB QMC solver since due to the minimization of the sign problem they lead to a significant computational speed up. A similar conclusion was reached in CT-HYB QMC study of Hubbard model\textsuperscript{69} where the Monte Carlo sign error was minimized by a canonical transformation that brought the hybridization matrix to a diagonal form.

In CI calculations, NOs lead to fewer bath orbitals - also resulting in an increased computational efficiency; however, here the savings are not so dramatic since including additional
bath orbitals in truncated CI schemes does not result in a significant cost increase.

It is important to showcase one more aspect necessary for a high accuracy quantum impurity solver for realistic problems. A quantum impurity solver capable of treating multiple regimes from weakly to strongly correlated ones, should be able to treat multiple partially occupied orbitals present in the impurity+bath system. To illustrate this statement, in Table III, we list impurity+bath occupations (8 impurity and 8 bath orbitals) for the H$_2$O chain in the cc-pVDZ basis (with a total of 50 orbitals). It is evident that for short interatomic distances this impurity+bath system has very few partial occupations, however for the stretched geometries, the number of partially occupied orbitals is significant. Consequently, if quantum chemistry methods are used as solvers for the impurity problems, they should be capable of treating many partially occupied orbitals - prompting us to conclude that single reference methods will generally not be capable of treating these impurity problems successfully.

### III. NUMERICAL ILLUSTRATIONS

Unless otherwise noted, the ORCA program$^{71}$ was used for all calculations using standard methods, e.g. FCI, CASSCF$^{72-74}$, and NEVPT2. The local modified version of the DALTON code$^{75}$ was employed to generate an RHF input necessary for GF2 and to evaluate (full) CI active space Green’s functions$^{30}$. Throughout this article, to compute potential energy curves accurately and to converge the electronic energy to $5 \times 10^{-4}$ a.u. with respect to the inverse temperature $\beta$ and the number of frequencies $w_{\text{max}}$, we employ different values of $\beta$ and $w_{\text{max}}$ along a single potential energy curve that yield a converged energy for different geometry points. We denote our method as SEET(CI/GF2)$[m \times n\alpha]$, where $m$ and $n$ are number of active spaces (impurity+bath problems) and number of orbitals in active spaces (number of impurity orbitals), respectively. The RASCI solver will be used when the impurity+bath problem is intractable with the FCI solver.

#### A. Non-local self-energy and screened interactions

First, let us show that SEET, where the non-local GF2 self-energy is included in the DMFT-like iterations, yields numerically very similar results to a DMFT procedure with screened interactions. As an example, we consider the H$_6$ chain in the STO-6G basis$^{68}$. All SEET calculations were performed in the SAO basis. In this work, we employ a statically screened impurity model, in which the screened interaction was obtained using GF2 and is defined via Eq. 20 at several chosen $\tau$ points. The effective impurity model present during the DMFT self-consistency is then solved using the FCI impurity solver.

The upper panel of Fig. 4 shows the on-site bare Coulomb interaction and screened interaction $U(\tau)$ at $\tau = \beta$ and $\beta/4.0$ as a function of the interatomic distance $R$. Note that the on-site bare Coulomb interaction in the SAO basis is not constant and changes with the distance. The on-site screened interaction should be equal to the bare interaction at large distances when the non-local interactions between atoms vanish. For short interatomic distances, due to the presence of large non-local interactions, the screened interaction is about twice smaller than the bare interaction.

The potential energy curves calculated with the on-site bare interaction, screened interactions, and SEET(FCI/GF2) are shown in the lower panel of Fig. 4. Note that SEET(FCI/HF)$-\nu_{\text{bare}}$, SEET(FCI/HF)$-U(\beta)$, and SEET(FCI/HF)$-U(\beta/4.0)$ corresponds to DMFT on top of HF with bare interactions $\nu_{\text{bare}}$, with on-site screened interaction $U(\beta)$, and with on-site screened interaction $U(\beta/4.0)$. The curves from FCI and SEET(FCI/GF2) with one site and two sites impurities are also provided for comparison.

In the next section, we will discuss the influence of the number of impurities on the potential energy curves and compare these with the SEET and DMFT results.

![FIG. 4. Upper panel: The on-site screened interaction $U(\tau)$, with $\tau = \beta$ and $\beta/4.0$, and bare Coulomb interaction as functions of the interatomic distance for H$_6$ chain in the STO-6G basis. Lower panel: Potential energy curves of H$_6$ chain in the STO-6G basis calculated using on-site bare and screened interactions. Results from FCI, SEET(FCI/GF2) with a single site and two sites impurities are also provided for comparison.](image-url)
TABLE I. Different values of the average sign and perturbation order (PO) for four possible impurities present in SAO and NO bases for H\textsubscript{6} molecule in the STO-3G basis set. Fig. 1 illustrates our impurity selection scheme using SAOs or Boys orbitals (spatial selection) and NOs (occupation selection). TRIQS\textsuperscript{70} program was used to perform CT-HYB calculations.

| \(\beta\) | Basis | Impurity A \(\langle\text{sign}\rangle\) | Impurity B \(\langle\text{PO}\rangle\) | Impurity C \(\langle\text{sign}\rangle\) | Impurity D \(\langle\text{PO}\rangle\) |
|---|---|---|---|---|---|
| 10 | SAO | 0.688 | 4 | 0.653 | 6 | 0.653 | 6 | 0.688 | 4 |
| 10 | NO | 0.845 | 0 | 0.705 | 1 | 0.478 | 3 | 0.500 | 2 |
| 50 | SAO | 0.232 | 16 | 0.262 | 28 | 0.262 | 28 | 0.232 | 16 |
| 50 | NO | 0.977 | 2 | 0.982 | 3 | 0.971 | 4 | 0.980 | 4 |
| 100 | SAO | 0.050 | 32 | 0.065 | 55 | 0.065 | 55 | 0.055 | 32 |
| 100 | NO | 0.914 | 4 | 0.946 | 6 | 0.970 | 8 | 0.978 | 9 |

TABLE II. Different values of the average sign and perturbation order (PO) for four possible impurities present in Boys and NO bases for H\textsubscript{6} molecule in 6-31G basis set. Fig. 1 illustrates our impurity selection scheme using SAOs or Boys orbitals (spatial selection) and NOs (occupation selection). TRIQS\textsuperscript{70} program was used to perform CT-HYB calculations.

| \(\beta\) | Basis | Impurity A \(\langle\text{sign}\rangle\) | Impurity B \(\langle\text{PO}\rangle\) | Impurity C \(\langle\text{sign}\rangle\) | Impurity D \(\langle\text{PO}\rangle\) |
|---|---|---|---|---|---|
| 10 | Boys | 0.652 | 3 | 0.730 | 5 | 0.730 | 5 | 0.652 | 3 |
| 10 | NO | 0.633 | 1 | 0.468 | 3 | 0.377 | 3 | 0.370 | 4 |
| 50 | Boys | 0.009 | 21 | 0.072 | 28 | 0.072 | 28 | 0.009 | 21 |
| 50 | NO | 0.919 | 2 | 0.922 | 3 | 0.931 | 5 | 0.964 | 8 |
| 100 | Boys | 0.000 | 42 | 0.004 | 56 | 0.004 | 56 | 0.000 | 42 |
| 100 | NO | 0.722 | 5 | 0.741 | 7 | 0.823 | 10 | 0.907 | 16 |

B. SEET outer loop: non-local self-energy relaxation

As we have mentioned in the SEET self-consistency description, SEET can be done as a one-shot procedure where \(\Sigma_{\text{non-local}}^{GF2}(\tau)\) coming from the initial self-consistent GF2 is not updated, or can be performed fully self-consistently where the non-local GF2 self-energy is updated during subsequent outer iterations involving GF2. Here, we would like to compare the one-shot and self-consistent procedures.

In systems where the entanglement between atoms is equivalent (for instance, hydrogen ring), separating the whole system into spatial fragments in the stretched regime will give rise to an incompatibility between local self-energy from high level theory and non-local self-energy from low level theory. Moreover, GF2 is not providing the self-energy that is accurate enough since the correlations are strong in the stretched regime. Consequently, in this regime, it is essential to carry out the outer iterations and update \(\Sigma_{\text{non-local}}^{GF2}(\tau)\) while starting from the zeroth order Green’s function that contains the strong correlation effects included due to the DMFT-like inner SEET iterations.

To reveal the importance of the non-local self-energy relaxation, especially, for the spatial fragment based embedding scheme, we plot the potential curves from SEET(FCI/GF2)[3×2o]SAO calculations for the H\textsubscript{6} ring and the H\textsubscript{6} chain in the STO-6G basis set. These results are summarized in Fig. 5.

Around the equilibrium, for both the H\textsubscript{6} ring and H\textsubscript{6} chain, one-shot and self-consistent GF2-DMFT schemes give almost identical results since the effect of non-local self-energy relaxation is negligible in the weakly correlated regime. When the interatomic distance is large, the one-shot and self-consistent curves significantly differ from each other and the former breaks down at the dissociation limit. A smooth dissociation curve is achieved when the GF2 non-local self-energy is re-
TABLE III. Orbital occupations from RASCI calculations for an impurity+bath problem consisting of 8 active (impurity) and 8 bath orbitals at \( R = 1.8 \) and 3.6 a.u. for the \( \text{H}_2\text{O} \) chain in the cc-pVDZ basis. The total number of electrons in the impurity+bath problems at \( R = 1.8 \) and 3.6 a.u. are 8 and 16, respectively.

| Orbitals | \( \text{R(H-H)} = 1.8 \) a.u. | \( \text{R(H-H)} = 3.6 \) a.u |
|----------|-------------------------------|-------------------------------|
| 1        | 1.9827                        | 2.0000                        |
| 2        | 1.9693                        | 2.0000                        |
| 3        | 1.9412                        | 2.0000                        |
| 4        | 1.8699                        | 2.0000                        |
| 5        | 0.1377                        | 1.8220                        |
| 6        | 0.0577                        | 1.7618                        |
| 7        | 0.0275                        | 1.6206                        |
| 8        | 0.0139                        | 1.3374                        |
| 9        | 0.0001                        | 0.6679                        |
| 10       | 0.0000                        | 0.3804                        |
| 11       | 0.0000                        | 0.2375                        |
| 12       | 0.0000                        | 0.1723                        |
| 13       | 0.0000                        | 0.0000                        |
| 14       | 0.0000                        | 0.0000                        |
| 15       | 0.0000                        | 0.0000                        |
| 16       | 0.0000                        | 0.0000                        |

laxed in the presence of the FCI local self-energy. Obviously, the effect of the non-local self-energy relaxation is more pronounced for the \( \text{H}_6 \) ring than for the \( \text{H}_6 \) chain since the entanglement between unit cells in the former is much stronger than that in the latter.

The relaxation of the non-local self-energy is further explored in Fig. 6, where we examine the case of \( \text{H}_6 \) ring. In principle, the non-local self-energy should decrease with the increasing inter-fragment distance since the non-local two-electron integrals \( v_{ijkl} \) where \( i, j, k, l \) belong to different fragments/orbitals should be vanishing. In practice, however, the non-local integrals do not decay fast enough giving rise to quantitatively inaccurate GF2 self-energy in the stretched atoms regime. Consequently, the non-local GF2 self-energy erroneously increases as a function of bond distance instead of decreasing, see Fig. 6. The gradual decrease can only be observed after employing the fully self-consistent SEET when the non-local GF2 self-energy is relaxed in the presence of the FCI self-energy accounting correctly for the strong correlations present at large bond distances. The fully self-consistent SEET yields a smooth transition from the weakly to strongly correlated regime as shown in Fig. 5.

C. Energy and spatial domain based embedding

The embedding procedure can be performed either in the energy or spatial domain. In the former the active orbitals are chosen as the most important NOs, whereas in the latter the whole system is split into different spatial fragments (SAOs/RNOs) that are physically/chemically meaningful.

At first, we consider the \( \text{H}_6 \) chain in the TZ basis set\(^7\) as shown in the upper panel of Fig. 7. An excellent agreement between SEET in NO basis and FCI curve can be observed when a full active space composed of six orbitals, SEET(FCI/GF2)\([6o]\)/NO, is used. Now, let us consider how splitting the full active space into different groups of NOs and RNOs (3\( \times \)2o means three groups of two orbitals) is influencing the results. At short distances, both SEET(FCI/GF2)\([3\times2o]\) in NO and RNO bases give very similar results. However, in the stretched regime, energies from the calculation in the RNO basis are much closer to the FCI curve than these from the calculation in the NO basis.

For the \( \text{H}_6 \) chain in the TZ basis, to get insight into the behavior of different orbital bases at dissociation, the occupations of six valence orbitals are plotted in the lower panel of Fig. 7. We compare the orbital occupations from SEET(FCI/GF2)\([3\times2o]\) in both RNO and NO bases to the exact ones calculated using FCI.

Note that there is a degeneracy of two pairs of active orbitals in the NO basis corresponding to two ends of the chain. FCI yields occupation numbers that smoothly transit from weakly to strongly correlated regime as the bond length increases. At short distances, i.e. \( R \leq 2.0 \) a.u., NO occupations are quite close to the FCI ones, while those in the RNO...
basis have slightly overestimated partial occupations. This difference can be understood in the following way. Since the correlation in the short bond regime is weak and the orbitals are mostly unoccupied or doubly occupied, SEET in the NO basis with the active space that is split into several groups is able to capture all the correlation effects well. In fact, even GF2 alone is good enough to describe correlation effects in this regime. However, in the short bond regime in the RNO basis the coupling between fragments is non-negligible, splitting the molecule into several fragments leads to missing the inter-fragment bonding, thus leading to less accurate results.

Upon bond stretching, fragments become nearly isolated and this is reflected by the degeneracy of all pairs of active orbitals in the RNO basis. Evidently, in the RNO basis the occupation numbers in the strongly correlated regime are almost parallel to the FCI ones. Thus, SEET in the RNO basis can yield a dissociation curve closely following the FCI one at long distances where the correlations are strong. In the NO basis, splitting the full active space into orbital groups is not sufficient to correctly describe the physics at dissociation limit. In this case, only the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) have reasonable occupations, while all the other orbitals have significantly different occupations from the FCI ones.

We now investigate a 2×4 hydrogen lattice, which is a more complicated example than the H₆ chain. The minimal STO-6G basis was used, so that dividing the whole system into spatial fragments can be done in the SAO basis. Fig. 8 shows the potential energy curves of 2×4 hydrogen lattice from SEET(FCI/GF2)[2×4o] in NO and SAO bases along with GF2 and FCI dissociation curves. For comparison, we also plot the curve from SEET(FCI/HF)[2×4o] in the SAO basis, i.e. non-local self-energy $\Sigma_{\text{non-local}}^{GF2}(i\omega)$ is not taken into account for this case. As mentioned previously, due to the lack of the self-energy terms describing non-local correlations, around equilibrium the energy from SEET(FCI/HF) with bare interactions is lower than FCI one, while in the intermediate regime the SEET(FCI/HF) energy is above the FCI one. Beyond the dissociation limit, the lattice is separated into isolated atoms, thus the electron interaction in such a system is actually just the on-site bare interaction. Therefore, in the dissociation limit, SEET(FCI/HF) gives energies in a good agreement with FCI results, however, the SEET(FCI/HF) curve is not smooth and has a discontinuity when a transition to the stretched regime happens. In the short bond region, energies from SEET(FCI/GF2)[2×4o] in the NO basis are better than GF2 energies and quite close to FCI ones, whereas SEET(FCI/GF2)[2×4o] in the SAO basis has significantly different occupations from the FCI ones.
D. Comparison with ground-state wavefunction methods

After discussing technical aspects of SEET, now we turn to showcase quantitative accuracy of SEET as compared to the standard quantum chemistry ground-state wavefunction methods. As mentioned previously, SEET(FCI/GF2) when a full active space is used can be directly compared to methods such as CASPT2 or NEVPT2 since in both of these methods, in an analogous manner to SEET, the strongly correlated (or active) orbitals are described by a higher level method than the weakly correlated (inactive) orbitals. In cases of CASPT2, NEVPT2, and SEET(FCI/GF2) the perturbative description involves a low level perturbative expansion. Therefore, it is necessary and interesting to make a numerical comparison between SEET(FCI/GF2) and CASPT2 or NEVPT2. We have done such comparison in our earlier work but only for SEET in the NO basis.

1. LiH$_{2}$ and LiH$_{4}$ chains in TZ basis

Here, we further analyze SEET results in the RNO basis. In Fig. 9, we present potential energy curves of LiH$_{2}$ and LiH$_{4}$ chains in the TZ basis. For the RNO basis, (LiH)$_n$ chain (with $n = 2, 4$) is divided into $n$ LiH fragments and the active orbitals that are used to construct the Anderson impurity models are constructed from two valence orbitals of each fragment. Both CASSCF and NEVPT2 correctly describe the dissociation. While GF2 yields accurate energies around the equilibrium, its curve is not parallel to the FCI one at long distances. However, when static correlation is properly treated using SEET(FCI/GF2), in both bases NO and RNO, the dissociation regime is described correctly. In particular for LiH$_{2}$, as seen in the upper panel of Fig. 10, although the SEET(FCI/GF2)[4o] in the NO basis yields a curve below the exact one, errors are of the same order as those of NEVPT2(4e,4o). SEET(FCI/GF2)[2x2o] in the RNO basis closely follows the NEVPT2 curve for all the distances considered here, thus having an error essentially equivalent to NEVPT2.

For LiH$_{4}$, we cannot evaluate FCI, thus we compare our results against NEVPT2. In the lower panel of Fig. 10, we plot the error in the energy per fragment for LiH$_{2}$ and LiH$_{4}$ molecules when compared to NEVPT2 energies per fragment. Interestingly, the difference between errors per fragment of LiH$_{2}$ and LiH$_{4}$ cases are very small ($\leq 1.0$ mHartree), indicating that the correct description of SEET(FCI/GF2) using LiH fragment as a repeating unit holds true regardless of the length of system.

2. H$_{50}$ chain in STO-6G basis

Now we test SEET(FCI/GF2) in the SAO basis on a well-known, non-trivial benchmark system, H$_{50}$ chain in the STO-6G basis. The exact solution is available from DMRG calculation. For a full comparison, we also present results from established theories capable of targeting strongly correlated molecules such as orbital-optimized antisymmetric product of 1-reference-orbital geminals (OO-AP1roG) and constrained-pairing mean-field theory combined with $\kappa$TPSSc functional [CPMFT($\kappa$TPSSc)]. Potential energy curves and errors relative to DMRG energies are displayed in the upper and lower panels of Fig. 11.
FIG. 10. Upper panel: LiH$_2$ chain in the TZ basis. Energy errors (in mHartree) $E_{\text{FCI}} - E_x$ for $X = \text{NEVPT2}(4e,4o)$ and SEET(FCI/GF2)[4o]/NO as a function of bond distances. Lower panel: Energy errors per fragment (in mHartree) $(E_{\text{NEVPT2}} - E_x)/N$ for SEET in the RNO basis for $N = 2, 4$ for both LiH$_2$ and LiH$_4$ in the TZ basis.

OO-AP1roG and CPMFT(κTPSSc) curves are far above and far below the DMRG reference, respectively. While the CPMFT(κTPSSc) curve displays huge non-parallelity errors near the equilibrium geometry, the OO-AP1roG curve remains nicely parallel to the DMRG curve. GF2 gives very good energies for short distances; however, it largely deviates from DMRG at long distances. SEET(FCI/GF2)[25x2o], where 25 impurity orbitals are embedded in the GF2 self-energy, yields a significantly improved energies at long distances when compared to GF2 alone. The errors can be further minimized when SEET(FCI/GF2)[5x4o+5x6o] calculation is carried out.

3. $H_{10}$ chain in cc-pVDZ basis

Here, we explore the concept of active space splitting where the full number of active orbitals is divided into several groups of orbitals used to build Anderson impurity models. To demonstrate that SEET is systematically improvable, when the number of orbitals used to build the impurities is increased, we performed calculations with different number of orbitals in the impurities for $H_{10}$ chain in the cc-pVDZ basis. The total number of orbitals in this basis set is 50 while the size of full active space is 10 orbitals. SEET results are summarized in Table IV along with GF2, NEVPT2(10e,10o) and DMRG energies for comparison. The DMRG data were computed using the BLOCK program. The errors relative to DMRG are shown in Fig 12 as a function of bond length. At short distances ($R < 2.0$ a.u.), GF2 energies are comparable to those from the NEVPT2(10e,10o) calculation. Upon bond stretching, when compared to DMRG, the GF2 error strongly increases, while NEVPT2(10e,10o) one slowly decreases. We can see that the errors of GF2 are significantly reduced when SEET(FCI/GF2)[2o+2x4o] calculation is carried out. For stretched distances, a systematic reduction of errors can be observed when the number of impurity orbitals is systematically enlarged starting from (2o+2x4o), (4o+6o), to (2o+8o). Interestingly, this error reduction with an increasing number of impurity orbitals is very systematic and is independent of the distances on the potential energy curve.

Consequently, one can expect that in cases where the number of active space orbitals is too large to be treated within one impurity, it is possible to split the active space orbitals among several impurities and systematically improve the answer. We
TABLE IV. Potential energies (in a.u.) as a function of bond distance (in a.u.) for H_{10} chain in the cc-pVDZ basis. Energies from GF2 and SEET with different number of impurity orbitals building the active space are compared to NEVPT2(10e,10o) and DMRG energies. FCI solver is used to treat [2o+2x4o] and [4o+6o] impurities, while RASCI solver is used for [2o+8o] impurity. Non-parallelity error [NPE] (a.u.) which is the difference between the largest and smallest errors with respect to DMRG references are also provided.

| R(H−H) | GF2 | SEET[CI/GF2] [2o+2x4o] | SEET[CI/GF2] [4o+6o] | SEET[CI/GF2] [2o+8o] | NEVPT2(10e,10o) | DMRG |
|--------|-----|------------------------|----------------------|---------------------|----------------|-------|
| 1.4    | -5.367 9 | -5.380 6 | -5.385 6 | -5.388 5 | -5.358 9 | -5.408 7 |
| 1.6    | -5.525 5 | -5.540 0 | -5.547 7 | -5.552 9 | -5.520 6 | -5.570 2 |
| 1.8    | -5.564 6 | -5.582 1 | -5.593 6 | -5.598 8 | -5.570 3 | -5.614 1 |
| 2.0    | -5.539 5 | -5.559 3 | -5.569 7 | -5.583 2 | -5.551 8 | -5.594 9 |
| 2.4    | -5.403 9 | -5.431 7 | -5.445 4 | -5.463 2 | -5.442 9 | -5.476 1 |
| 2.8    | -5.235 2 | -5.271 4 | -5.292 1 | -5.320 4 | -5.305 7 | -5.334 4 |
| 3.2    | -5.074 1 | -5.119 3 | -5.148 7 | -5.192 9 | -5.185 3 | -5.212 5 |
| 3.6    | -4.935 3 | -4.995 0 | -5.032 8 | -5.082 2 | -5.095 4 | -5.123 8 |
| NPE    | 0.147 7 | 0.100 8 | 0.068 6 | 0.028 8 | 0.021 5 |

FIG. 12. Errors in energy (in a.u.) with respect to DMRG reference as a function of R(H−H) for GF2, SEET with different active spaces in the NO basis, and NEVPT2(10e,10o). All energies (in a.u.) are presented in Table IV.

would like to stress that this systematic improvement will become crucial for systems where the exact answer is unknown, thus the only way of assessing if the level of accuracy given by SEET is sufficient will be coming from internal SEET criteria and checking if the answer obtained does not change drastically upon enlarging the number of impurity orbitals.

IV. CONCLUSIONS

In this paper, we have presented a detailed discussion of the molecular Green’s function quantum embedding scheme called SEET. The self-energy separation characteristic for SEET onto strongly correlated/active/subsystem and weakly correlated/inactive/environment parts is completely general and does not specify how the self-energies for these fragments will be evaluated in practice. While many schemes are possible, in this paper we used a scheme where first the whole molecule is treated by the perturbative self-consistent GF2 approach and then selected strongly correlated orbitals are used to build impurity+bath models that are solved in DMFT-like procedure in the presence of self-energy coming from the weakly correlated/inactive/environment orbitals.

We aim for SEET to be systematically improvable, without empirical parameters, and reaching chemical accuracy. Consequently, we discussed many aspects of SEET that were developed by us to fulfill these strict demands. SEET is a Green’s function method capable not only of delivering ground state energies but also many more physically relevant quantities such as free energies, ionization potentials (IP) and electron affinities (EA), or temperature-dependent magnetic susceptibility; however, here we focused on analyzing SEET results for small molecular examples where multiple ground state methods are known to give excellent results and ground state energies can be easily used to assess the SEET performance.

We started our considerations by explaining the double self-consistency loop present in SEET, where in the inner DMFT-like loop the active space/impurity self-energy is updated using an accurate many-body solver. The outer loop requires an update of the self-energy for inactive orbitals performed by us at the GF2 level; however other inexpensive ab initio methods could also be used for the calculation of the inactive orbitals self-energy. We observed, when analyzing numerical results, that this outer loop’s self-energy update is crucial for classes of systems where the initial perturbative description was not quantitative.

Next, we have analyzed different schemes for selecting the strongly correlated/active/subsystem orbitals either based on the energy (occupations) or spatial criteria. We stress that in the energy (occupation) scheme the strongly correlated orbital selection is done mainly based on the occupations of correlated one-body density matrix, thus not only relying on intuitive means. Moreover, in the result section for H_{10} chain, we have demonstrated that the results of such a selection scheme
can be systematically improved. We have also analyzed how different orbital bases, SAO, RNO, and NO influence SEET results. As expected, we found that NO basis, which allows us easily describe a large inter-orbital entanglement, yields very good results for equilibrium geometries; however, in the limit of separated fragment bases that localize orbitals, such as SAOs/RNOs, yield better energies.

We discussed optimal impurity solvers that can be used with SEET, stressing that a full realistic Hamiltonian should be employed for strongly correlated orbitals if one desires systematically improvable and accurate results. Moreover, we have observed that suitable solvers (to be applicable in multiple regimes from weakly to strongly correlated) must be able to deal with near degeneracy and have to be able to treat multiple strongly correlated impurity orbitals. Finally, we highlighted the influence of the orbital basis onto the performance of impurity solvers by noticing that for molecular examples the NO basis seems to be particularly advantageous for the CT-HYB and RASCI solvers.

In SEET the influence of the non-local interactions on the strongly correlated/active/embedded system is contained in the non-local self-energy evaluated at the perturbative level. We have compared SEET results with DMFT results where the non-local self-energy evaluated at the perturbative level allows us to avoid the downfolding procedure and construct the presence of self-energy coming from the inactive orbitals. Coulomb interactions present in the impurity model solved in the Green’s function analog of CASPT2 or NEVPT2 procedure.

Finally, we have performed multiple total energy comparisons against established quantum chemistry methods such as NEVPT2 and DMRG. For calculations in the NO basis and when a full active space is used, SEET can be considered as a Green’s function analog of CASPT2 or NEVPT2 procedure. We have demonstrated that when employing a full set of active space orbitals SEET(CI/GF2) yields energies that are very close to NEVPT2. However, unlike CASPT2 or NEVPT2, SEET(CI/GF2) is a perturb and diagonalize scheme that does not require storing or evaluating of two-, three-, or four-body reduced density matrices and avoids the intruder state problem present in CASPT2. Here, we focused on analyzing how splitting of the full set of active orbitals into several impurities can be done in a systematic manner leading to an improvable and controlled procedure. We have shown that both in the spatial and energy domains, SEET results can be systematically improved by using multi-site impurities. We analyzed the spatial domain using the example of H$_{10}$ chain in STO-6G basis, while the energy domain was examined on the example of H$_{10}$ chain in cc-pVDZ basis. Moreover, we have shown that SEET results in a small non-parallelity error when compared against DMRG and other established multi-reference methods. These results that indicate systematic trends provide us with a number of important self-contained assessment tools that in the future can be used to help us rigorously analyze the accuracy of SEET in the absence of known results.

We believe that the considerations and results presented here are further establishing SEET as a quantum embedding Green’s function method that is controlled, systematically improvable, and not only able to reach accuracy comparable to currently established active space quantum chemistry methods but also flexible enough to yield other spectral and thermodynamic quantities than the ground state energies.

ACKNOWLEDGMENTS

T.N.L., A.A.K, and D.Z. would like to acknowledge support from the US Department of Energy (DOE) grant No. ER16391. Authors also thank Ushnish Ray and Garnet Kin-Lic Chan for providing the DMRG data for the H$_{10}$ chain in the cc-pVDZ basis set.

---

1. latran@umich.edu
2. On leave from: Ho Chi Minh City Institute of Physics, VAST, Ho Chi Minh City, Vietnam
3. G. Ghigo, B. O. Roos, and P. A. Malmqvist, Chem. Phys. Lett. 396, 142 (2004).
4. K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. 94, 5483 (1990).
5. C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, J. Chem. Phys. 114, 10252 (2001).
6. C. Angeli, R. Cimiraglia, and J.-P. Malrieu, J. Chem. Phys. 117, 9138 (2002).
7. R. J. Bartlett and M. Musial, Rev. Mod. Phys. 79, 291 (2007).
8. B. Jeziorski and H. J. Monkhorst, Phys. Rev. A 24, 1668 (1981).
9. C. Angeli, R. Cimiraglia, and J.-P. Malrieu, J. Chem. Phys. 134, 114108 (2011).
10. I. Lindgren, Int. J. Quant. Chem. 14, 33 (1978).
11. D. Mukherjee, R. K. Moitra, and A. Mukhopadhyay, Mol. Phys. 30, 1861 (1975).
12. P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, Chem. Rev. 112, 108 (2012).
13. J. Gräfenstein and D. Cremer, Chem. Phys. Lett. 316, 569 (2000).
14. J. Gräfenstein and D. Cremer, Mol. Phys. 103, 279 (2005).
15. R. Pollet, A. Savin, T. Leininger, and H. Stoll, J. Chem. Phys. 116, 1250 (2002).
16. G. L. Manni, R. K. Carlson, S. Luo, D. Ma, J. Olsen, D. G. Truhlar, and L. Gagliardi, J. Chem. Theory Comput. 10, 3669 (2014).
17. R. K. Carlson, D. G. Truhlar, and L. Gagliardi, J. Chem. Theory Comput. 11, 4077 (2015).
18. A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. 68, 13 (1996).
19. G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Rev. Mod. Phys. 78, 865 (2006).
20. T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, Rev. Mod. Phys. 77, 1027 (2005).
21. L. Hedin, Phys. Rev. 139, A796 (1965).
22. S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. 90, 086402 (2003).
80 T. H. Dunning, J. Chem. Phys. 90, 1007 (1989).
81 U. Ray and G. K.-L. Chan, (2016), private communication.
82 S. Sharma and G. K.-L. Chan, J. Chem. Phys. 136, 124121 (2012).
83 R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, J. Chem. Phys. 142, 034102 (2015).